HANDBOOK OF PUBLIC FINANCE

Edited by

Jürgen G. Backhaus
Richard E. Wagner

Handbook
of
Public Finance

General Editors: **Juergen G. Backhaus**
University of Erfurt
Richard E. Wagner
George Mason University

Contributors:

Andy H. Barnett
Charles B. Blankart
Thomas E. Borcherding
Rainald Borck
Geoffrey Brennan
Giuseppe Eusepi
J. Stephen Ferris
Fred E. Folvary
Andrea Garzoni
Heinz Grossekettler
Walter Hettich
Scott Hinds
Randall G. Holcombe
Jean-Michel Josselin
Carla Marchese
Alain Marciano
William S. Peirce
Nicholas Sanchez
David Schap
A. Allan Schmid
Russell S. Sobel
Stanley L. Winer
Bruce Yandle

Handbook
of
Public Finance

Editors

Jürgen G. Backhaus
University of Erfurt

and

Richard E. Wagner
George Mason University

Visit Springer's eBookstore at: http://ebooks.kluweronline.com
and the Springer Global Website Online at: http://www.springeronline.com

Handbook of Public Finance

Table of Contents

Chapter 1

SOCIETY, STATE, AND PUBLIC FINANCE: SETTING THE ANALYTICAL STAGE

Jürgen G. Backhaus
University of Erfurt
juergen.backhaus@uni-erfurt.de

Richard E. Wagner
George Mason University
rwagner@gmu.edu

Abstract	Much of contemporary public finance can be described as being either Wicksellian or Edgeworthian in character. In the former, fiscal phenomena arise through complex processes of exchange; in the latter, they arise through acts of choice by some maximizing agent. This chapter traces these alternative orientations in the history of thought as far back as the Cameralists, compares these writers with Adam Smith, and contrasts the Wicksellian and Edgeworthian orientations toward public finance.
Keywords:	Adam Smith, cameralism, Francis Edgeworth, Knut Wicksell, choice-theoretic public finance, catallactic public finance
JEL classification:	B10, H40

The essays that comprise this *Handbook* cover a wide variety of topics in the theory of public finance. As a field of systematic academic inquiry, public finance arose before economics or political economy. For instance, more than 90 chairs in public finance had been established in Europe before the first chair was established in political economy (Backhaus, 2002, p. 615). The first scholars of public finance were the Cameralists, who emerged in central Europe in the 16th century. For a long time after its Cameralist founding, public finance was conceived as a multi-disciplinary field of study, and most certainly not simply a subset of economic theory. The object that public finance scholarship examined, the public household, was examined in a manner that sought to in-

tegrate the economic, political, legal, and administrative elements of public finance.

In the post-war period, a splintering of approaches has taken place, sufficiently so that it is meaningful to distinguish between the old public finance and the new, recognizing that new is not a synonym for better. The old public finance has not been replaced by the new. Both are practiced currently, and are competing for the affection and attention of scholars. If the old-style public finance is still called public finance, the new style is often called public economics instead. This distinction is found in contemporary texts and journals. For instance, there is a *Journal of Public Finance and Public Choice* and a *Public Finance Review.* There is also a *Journal of Public Economics* and a *Journal of Public Economic Theory.*

The distinction between what might be called old-style and new-style public finance was recognized clearly in Richard Goode's (1970) commentary on the theory of public finance. There, Goode compared the treatment of public finance in two different social science encyclopedias, written a generation apart. One of these was the *International Encyclopedia of the Social Sciences,* which was published in 1968. The other was the *Encyclopedia of the Social Sciences,* which had been published in 1930. While Goode duly noted the theoretical advances that had occurred in economics between 1930 and 1968, he also lamented the narrowing of the subject matter of public finance. Goode concluded his lamentation on the state of public finance by asserting that "a sophisticated and unified treatment of the economic, political, legal, and administrative elements of public finance is needed. Unification would represent a return to a tradition as old as that of the cameralists, but for modern readers sophistication can be attained only by rethinking old problems and using new techniques. There is much to be done and work for a variety of talents" (p. 34).

The difference that Goode noted maps nicely into the distinction we have advanced between an old-style public finance and a new-style, noting again that old and new refer simply to the times when those approaches arose and represent no judgment about relative merit or quality. As we do not embrace a Whig theory of history, we do not regard the new style of public finance as *ipso facto* superior to the old style. Indeed, there is much about the old-style that we regard as superior to the new. The essays in this *Handbook* generally affirm the orientation toward public finance that informed both Goode's appraisal and his assessment about the potential value of scholarly inquiry that sought more fully to integrate the economic, social, political, legal, and administrative aspects of public finance.

This opening essay is *not* a survey of the historical development of fiscal theory. It would take a very large book indeed to accomplish this. Our intention here is simply to provide some elementary historical orientation toward the

two styles of public finance. We do this by referring to two main theoretical antinomies that run throughout the history of public finance, and to do so with reference to a few of the main historical contributors to the theory of public finance.

The first antinomy we explore is between conceptualizing the state as *intervening into* the economic order and the state as *participating within* the economic order. To do this we return briefly to consider the Cameralist origins of the theory of public finance, and to contrast the Cameralist approach with that of their classical British contemporaries, particularly Adam Smith. The second antinomy is between treating the state as some maximizing agent and treating the state as an institutional framework within which myriad individual agents interact. The primary historical figures who represent this second antinomy are the British economist Francis Edgeworth (1897) and the Swedish economist Knut Wicksell (1896). For Edgeworth, public finance was a choice-theoretic enterprise, with the policy choices of a state being assimilated to the market choices of an individual. In sharp contrast, Wicksell treated public finance as a catallactic enterprise, where the state provided an institutional framework within which individuals with differing values and preferences interacted. The extent to which those interactions proved generally beneficial to everyone, as against providing gains to some and losses to others, depended on the institutional framework within which political participants interacted with one another.

These two antinomies can in principle generate a four-fold description of analytical possibilities. A model of an interventionist state can in principle be developed within either a choice-theoretic or a catallactic framework. Similarly, a model of a participative state can be developed within either a choice-theoretic or a catallactic framework. For the most part, however, the interventionist state has been assimilated to a choice-theoretic framework, and with the participative state being assimilated to a catallactic framework. In the remainder of this opening essay, we will first illustrate the intervention-participation antinomy as it is found in the formulations of the Cameralists and the British Classicals. Then we use Edgeworth and Wicksell to illustrate the antinomy between the choice-theoretic and the catallactical approaches to public finance. We close this essay by briefly describing the essays that comprise this *Handbook*.

1.   THE CAMERALIST ORIGINS OF PUBLIC FINANCE

The cameralists emerged around 1500, and were mostly located in the German-speaking lands. By the time they had disappeared by the middle of the 19th century, they had amassed a collective bibliography of more than 14,000

items, according to Magdalene Humpert (1937). Joseph Schumpeter (1954, pp. 143-208) described the cameralists well when he referred to them as "Consultant Administrators." They were both consultants and administrators. They were consultants to the various kings, princes, and other royal personages who ruled throughout those lands. Indeed, the term cameralist derives from camera or kammer, and refers to the room or chamber where the councellors to the king or prince gathered to do their work. The cameralists were not, however, anything like contemporary academic consultants. They were real-world administrators as well. They were engaged in such activities as managing mines and glass works. Many of the cameralists also held academic posts. The first chairs of cameral science were established in 1727, in Halle and Frankfurt on the Oder, and by the end of the 18th century 23 such chairs had been established (Backhaus, 1993). The cameralists were partly economists, partly political scientists, partly public administrators, and partly lawyers. They approached their subject matter in a manner that used all of these talents and capacities.

The Peace of Westphalia in 1648 recognized more than 300 independent units of governance within the cameralist lands. Cameralism arose under conditions of high political fragmentation. A cameralist land faced different circumstances than the contemporaneous mercantist regimes. There was no concern within the cameralist lands about influencing terms of trade, about the use of colonies as instruments of policy, and about one's relative standing among the preponderant powers. All of these concerns were foreclosed by circumstance to those who ruled within the cameralist lands. The focal point of cameralist concern was on survival of the regime. Survival, in turn, required a military capacity. It also required economic development, which in turn required the acquisition of advanced technologies, the improvement of human capital within the population, the creation of new enterprises, and the growth of population with a view to creating human capital.

This concern about development took place within regimes that were both absolutist and severely constrained. The prince was the ruler of his lands. He did not have to worry about surviving periodic elections, and he could hope to pass his principality along to his eldest son. His ability to do this, however, varied directly with the extent of economic progress within his land. A prince whose land was supporting a growing population of energetic and enterprising subjects would both be wealthier and face better survival prospects than a prince of a land where the population was stagnant or declining, and whose subjects were dull and lethargic or else departing the country.

If one were to construct a model of the cameralist vision of the state, it would resemble a model of a business firm. The state's lands were potential sources of revenue. Forests could be harvested, game could be caught, and mines could be built and worked. The ruler would also sponsor an assortment

of commercial enterprises, including such things as the operation of a glass-works or a brewery. Taxes occupied a secondary position as a source of revenue. They were a last resort option for public finance, and not the first source of revenue.

The cameralists' general predisposition against taxation as an instrument of public finance reflects an orientation that the state acts as a participant within the economic order. Individuals had their property and the state had its property. The state should be able to use its property to generate the revenues required to finance its activities. Or at least those enterprise revenues should support the major portion of state activity. Some of the cameralists argued that taxes should be earmarked for the support of the military, while all activities concerned with internal development should be financed from the prince's net commercial revenues. In any case, the state's enterprises were to be the primary source of revenue for the state. It was understood that the state would have significant expenses associated with its activities. These expenses, however, were not to become drains upon the private means of subjects. They were to be met from the lands and enterprises that constituted the state's property.

It is instructive to compare the approach to taxation taken by Johann Heinrich Gottlob von Justi (1771) and Adam Smith (1776), particularly with respect to the limits placed on the use of the power to tax. Smith, of course, is one of the premier figures of classical liberalism, and it is hardly surprising that his maxims of taxation are widely thought to serve as strong limits on the power to tax. Smith's four maxims of taxation have been stated repeatedly in public finance texts since he first articulated them in 1776. These are:

(1) Taxes should be levied in proportion to property.
(2) Taxes should be certain and not arbitrary.
(3) A tax should be convenient to pay.
(4) A tax should be economical to administer, for both the taxpayer and the state.

Justi (1771, pp. 549-565) similarly articulates maxims for taxation, though these maxims, unlike Smith's, have not been carried forward in the public finance literature. Justi's maxims actually go beyond Smith in limiting the power to tax. In addition to the limits articulated by Smith, Justi's maxims for taxation included requirements that taxes should never cause taxpayers to reduce their capital, harm their welfare, or violate their civil liberties.

While Justi placed even stronger limits on the power to tax than did Smith, even more notable is the sharp difference they accorded to taxation within the overall scheme of public finance. Smith regarded taxation as the primary source of public financing, and thought ideally that it should be the sole source of public finance. For instance, Smith preceded his presentation of tax maxims with an argument that the state should eliminate its property and the revenues

derived therefrom. In sharp contrast, Justi preceded his discussion of tax maxims with a discussion of why taxation should be a last resort or secondary means of public finance. Indeed, Justi argued that ideally the state would not tax at all, and would derive all of its revenues from its enterprises. In sharp contrast, Smith argued that ideally the state would own no enterprises, and would derive all of its revenue from taxation.

This difference between Justi and Smith reflects one of the important orienting principles of the cameralists, namely, that the state acts as one participant among many within the society and its economic order. The cameralist advice on the use of state budgets and other policy instruments to promote the happiness of the state and its subjects took place within a presumption that the state itself was located inside the economic order and not outside it. Civil society and the state are nonseparable and co-emergent, and the state participates within the economy on the same terms as all other participants. The state finances its activities through market activities just as everyone else does; it acts within the market economy. By contrast, for Smith the state intervenes into civil society and its processes. Everyone else in society acts within the market, and the state then intervenes into the market by taxing various results of market activity.

Geoffrey Brennan and James Buchanan (1980) construe the state as a revenue-maximizing beast, a leviathan. While the leviathan of the Bible lived in the sea, it is easy enough to imagine it as living on the land. Smith's maxims for taxation are a recipe for living with the leviathan by doing such things as clipping the beast's nails and filing its teeth. A beast it will always be, and the objective of tax maxims should be to limit the damage the beast causes. Justi's maxims for about the primacy of enterprise revenues and taxation as a last resort measure represent a contrary intellectual orientation that would seek to domesticate the beast. Whether it is actually possible, or the extent to which it is possible to domesticate the beast is a different matter that has occupied a good number of scholars, fiscal and otherwise.

Cameralistic public finance was a choice-theoretic approach to public finance. The phenomena of public finance, state revenues and expenditures, arise out of a ruler's optimizing choices. It is quite different in modern democratic regimes. The phenomena of public finance do not arise from someone's optimizing choice, but rather arise through interaction among the many participants within the fiscal process. This interactive or catallactic approach to public finance leads often to quite different implications for public finance than the choice-theoretic approach (Wagner, 1997, 2002). The dominant portion of contemporary public finance has maintained the choice-theoretic orientation toward public finance, as if fiscal phenomena are still generated through the same processes that were in place in mercantalistic and cameralistic times. This astonishing situation was noted in 1896 by Knut Wicksell (1958, p. 82),

when he complained that the theory of public finance "seems to have retained the assumptions of its infancy, in the seventeenth and eighteenth centuries, when absolute power ruled almost all Europe."

2.   FROM CAMERALIST ORIGINS TO CONTEMPORARY PUBLIC FINANCE

Cameralist public finance ended early in the 19th century, giving way to its closely related analytical cousin, Staatswissenschaften. Perhaps the best known treatise in the tradition of Staatswissenschaften is Adolf Wagner's (1883) two volume set. The contemporary relevance of this analytical tradition is probed, moreover, in a symposium that appeared in the September 2001 issue of the *European Journal of Law and Economics.* The tradition of Staatswissenschaften fits clearly within the framework of the older style of public finance, as does the more widely known contribution of Knut Wicksell. The primary fault line that divides the older style from the newer style is the same fault line that Wicksell articulated, and concerns the political setting within which theorizing about public finance occurs. The newer-style public finance treats the political setting as one of old-style absolutisms where fiscal phenomena can be assimilated to some ruler's maximizing or optimizing choices. This is the type of public finance to which Wicksell objected more than a century ago. The older style public finance, in its contemporary incarnation, deals seriously with Wicksell's objection, by developing an interactive or catallactical approach to public finance. For contemporary times, the seminal articulation of the choice-theoretic orientation can be attributed to the British economist Francis Edgeworth (1897), while the seminal articulation of the catallactic orientation can be attributed to the Swedish economist Knut Wicksell (1896).

Cameralist public finance was choice-theoretic, in that it addressed the fiscal choices of some ruler. The cameralist regimes were autocratic. Fiscal programs were the choices of a ruler. It is reasonable to assimilate a theory of cameralist public finance to a model of the optimizing choices of a business firm. Indeed, the prince as a firm was a central part of the cameralist analytical framework. Edgeworthian public finance is likewise choice theoretic, only the ruler to which the formulations are addressed is some fictional being, at various times characterized as a social welfare function or a benevolent despot. Wicksellian public finance is catallactic and not choice-theoretic in its orientation toward fiscal phenomena. It also construes the state as one participant among many within the economic processes and activities of a society. In this construction of the state as a participant, the Wicksellian tradition contrasts markedly with the Edgeworthian tradition and its focus on the state as intervening into the economic process.

2.1.   The Edgeworthian, Choice-Theoretic Tradition

This tradition treats public finance as the study of government intervention into the economy, typically to maximize some notion of social welfare. Edgeworth raised the question of how a government would impose taxes within a nation if it wanted to raise those taxes with a minimum amount of sacrifice to taxpayers. For a given amount of revenue to be raised, Edgeworth's ideal state would be one that imposed the least amount of sacrifice upon taxpayers in raising its revenue.

Taxes would reduce the disposable income of taxpayers, and the sacrifice that would be involved would depend on the utility that the taxpayer derived from that sacrificed income, in Edgeworth's utilitarian formulation. With taxpayer sacrifice being measured by lost utility, the total amount of sacrifices created by a particular level of taxation required an ability to sum sacrifices across taxpayers. If it is presumed that the relation between income and the utility provided by that income is identical for all taxpayers, and if it is further presumed that the marginal utility of income declines with income, a simple system of taxation arises, as a first approximation. The least sacrifice of utility comes from the highest income in society. Hence, taxation should take away the highest income first, if it is to minimize the total amount of sacrifice from taxation. What results is a tax that pares down incomes from the top until the required amount of revenue is raised.

To be sure, Edgeworth himself noted that this was only a first approximation, because the effort to impose such a punitive, 100 percent marginal rate of tax on high incomes would eliminate the incentive to earn those incomes, which in turn would make other people worse off. This insight was later formalized in what has become known as the theory of optimal taxation, inspired by Frank Ramsey (1927) and surveyed in James Mirrlees (1994). This theory fits within the Edgeworthian, choice-theoretic tradition, in that it construes the state as facing its own problem of utility maximization. In this case, the state is to maximize aggregate social utility, though its ability to do this is constrained by the recognition that high taxes will reduce the willingness of people to earn income. Where Edgeworth would apply a 100 percent rate of tax to the highest dollar of income in a society, the theory of optimal taxation would let that last dollar be free from tax.

Some of the details involved in these formulations will be addressed in some of the later essays. All we would note here is that this Edgeworthian, choice-theoretic approach to public finance treats the phenomena of public finance as arising from the maximizing choices of a benevolent entity, the state. The state stands outside the market economy and its participants. The people who participate in the market economy may write the first draft, so to speak, but it is the state that revises and perfects the manuscript.

2.2.   The Wicksellian, Catallactical Tradition

Where the Edgeworthian, choice theoretic tradition is one where public finance is viewed as the means by which the state intervenes into the market economy to perfect its results, the Wicksellian, catallactical tradition views public finance as the study of how people participate through government to achieve their various ends. The state does not stand above the market economy and its participants. The same people who participate in the market economy participate in state governance as well.

Fiscal phenomena are not the product of some ruler's maximizing choices, but rather emerge through interaction among people. This interaction might be beneficial for everyone or nearly everyone, or it might be beneficial for only a few, and costly for many others. The state is treated as a nexus of contractual and exploitive relationships. The extent to which those relationships are contractual or exploitive depends, as some of the following essays will explore, on the constitutive structure of governance that is in place.

As a matter of general principle, political relationships are both contractual and exploitive. It is fine to say that taxes are the prices we pay for civilization. This doesn't mean, however, that the relationship between citizens and state is the same as the relationship between customers and the retail outlets they frequent. A customer can refuse to buy and, moreover, generally can return merchandise that turns out to be defective or otherwise unsatisfactory. There is no option to do this in politics. To say that civilization is being priced too highly and to withhold payment will only land the protester in prison. And there is certainly no point in asking for a refund by claiming that the state's offerings weren't as good as its advertisements claimed them to be.

To speak of a catallactical approach to public finance is not to claim that the phenomena of public finance arise through voluntary interaction among people. It is only to say that those phenomena arise through interaction among people, the very same people as who interact with one another within the market economy. Much of the phenomena of public finance surely arise through duress and not through genuine agreement. This aspect of duress was given particular stress in a good deal of the Italian scholarship on public finance, and which is surveyed in James Buchanan (1960).

The sometimes sharp differences between the older and newer styles of public finance is revealed in the dueling book reviews that accompanied the publication in 1934 of Antonio De Viti De Marco's treatise, *Principii di Economia Finanziaria.* De Viti's treatise was a major statement of the Italian tradition of public finance, a tradition that arose within an interactive framework. It was reviewed in the August 1934 issue of Economica by Fredric Benham, who asserted that De Viti's book "is probably the best treatise on the theory of public finance ever written." Benham laments that sorry state of public finance in

England, which was dominated by the choice-theoretic orientation that Wicksell decried, could be improved greatly through a strong infusion of De Viti's orientation. Benham also noted the strong complementarity between the approaches taken by De Viti and Wicksell.

In sharp contrast, Henry C. Simons reviewed the English translation of De Viti that appeared in 1937, in the October 1937 issue of the *Journal of Political Economy*. Simons began by observing that "the Italian literature of public finance has long been held in high esteem; but its claims to distinction have rested mainly upon works which have been inaccessible to those of us who lacked facility with the language. The translations [both German and English translations were being reviewed by Simons] of De Viti's famous treatise are thus doubly welcome, for they will make possible a more informed consensus, both as to the merits of Italian economics and as to competence of the interpretation and appraisal which it has received in other countries."

After describing this initial sense of eager anticipation, Simons offered his judgment: "Careful reading … has left the reviewer with no little resentment toward the critics who induced him to search in this treatise for the profound analysis and penetrating insights which it does not contain. The *Principii* is revealed to him, not as a great book, but as a … monument to … confusion." Simons continued by asserting that "there is not a single section or chapter which the reviewer could conscientiously recommend to the competent student searching for genuine insights and understanding."

Simons concludes by taking on Benham's review three years earlier: "If his book is 'the best treatise on the theory of public finance ever written,' one hopes that it may be the last.… To say that it is distinguished among treatises in its field is to praise it justly and, at the same time, to comment bitterly on the quality of economic thought in one of its important branches. To call it a great book, however, is a disservice to the cause of higher standards and better orientation in economic inquiry."

That two reviewers, each so prominent in his time, could be so opposed in their appraisals can only testify to a sharp clash in the presumed domains of fiscal inquiry. De Viti and Benham shared an orientation toward the domain of public finance that was antagonistic with Simon's orientation. This clash of orientations toward public finance, moreover, took shape in the late 1800s and has been carried forward to this day

Within the Wicksellian, catallactical tradition, primacy of analytical attention is placed upon the institutions of governance, both market governance and, especially, political governance. This contrasts sharply with the Edgeworthian, choice-theoretic tradition where the primary analytical attention is placed upon prices and resource allocations. The essays in this *Handbook* are generally written from within a broadly Wicksellian or catallactical orientation toward public finance. The state is not treated as some exogenous force

that perfects and corrects the outcomes of the market economy. The actual fiscal conduct of the state emerges through complex interactions among fiscal and political participants, and the precise character of those interactions is constrained and shaped by a governing institutional and constitutional framework.

3.   THE PRESENT RELEVANCE OF PAST FORMULATIONS

A choice-theoretic approach to public finance was suitable in cameralist and mercantilist times. A cameralist ruler could reasonably be described as seeking to use his fiscal means to promote his dynastic ends. For the cameralists it was historically accurate to ascribe the phenomena of public finance to the choices of the rulers. The state's revenues depended on the ruler's choices about how to operate his mines and forests and how to farm his lands. The extent to which state expenditures were directed to projects that might increase future productivity were likewise objects of choice for the ruler. Suppose two kingdoms were observed to undertake different expenditure programs. In the first kingdom expenditures were heavily oriented toward such investments as draining swamps and building roads that would be likely to increase future production. The budget in the second kingdom, however, did little about swamps and roads, and instead spent lavishly on amusements for the king and his court. It would be reasonable in this case to compare the budgetary choice of the two kingdoms, and to say that the first king had a lower time preference, or was otherwise more far-sighted than the second king. To the extent it is possible to make inferences about preferences from the observation of choices with respect to private choices, it would be possible to do the same thing with respect to state choices within the cameralist setting. To be sure, the conduct of cameralist rulers was relatively civilized, and nothing like the experience with dictators in the 20th century. The conceptual construction of a benevolent despot perhaps finds historical validation in the cameralist period. That does not, however, render empirically valid the use of constructions based on benevolent despots in public finance today.

Whether budgets in a democratic regime are tilted toward amusements or capital projects would not be a source of information about some person's preferences. Budgets emerge out of interactions among participants, and those interactions are governed and shaped by a variety of procedural rules. The people who participate in a market make their various choices, but it makes no sense to speak of the market itself as making choices. The market simply registers and reflects the choices and interactions among the participants. It is the same with budgetary outcomes within a democracy. Furthermore, the same set of people can generate quite different budgetary outcomes, depending on the institutional framework within which the budgetary process proceeds. In this

respect, there is an indefinite number of particular budgetary processes that can be imagined, and it is conceivable that a wide variety of budgetary outcomes could be generated, if the experiment were performed of having the same people engage in successive interactions across differing institutional frameworks. This consideration suggests immediately that a post-cameralist public finance would place particular importance and significance on the institutional framework within which fiscal outcomes emerge.

The cameralists were clearly agents for their royal principals. Principals who were unhappy with their cameralist agents would dismiss them, and could well imprison them for malfeasance. Justi, for instance, died while imprisoned for alleged financial mismanagement. While modern democracies are quite different from the cameralist absolutisms, such categories as principal, agent, and property are present now just as they were then. The cameralists spoke of subjects. We now speak of citizens. It is the citizens who are the principals in a democracy. The head of state was the principal in cameralist times, but is now the agent. The same relationship of agency exists in modern democracies as existed in cameralist times, only the substantive character of that relationship is different in many respects.

All agency relationships raise questions of how strongly the agent will promote the desires of the principals. Shopping centers, apartment complexes, and hotels all provide state-like services in a cameralist-like setting (see, for instance, MacCallum, 1970 and Foldvary, 1994). What these organizations do is offer forms of tie-in sales, where private and public services are offered as a package. Apartments and hotels offer rooms to residents. The rental price, however, also finances the provision of an array of public services. Hotels will have subways that run vertically. Hotels usually sweep their streets daily. Hotels and apartment complexes typically provide a variety of parks and playgrounds. Walt Disney World in Florida offers the same kind of arrangement, only it covers 45 square miles of territory. All topics relating to property and agency within the conduct of government would fit naturally within a post-cameralist orientation toward public finance.

From the perspective of today, we would call the cameralists multidisciplinary, with the primary disciplines being economics, politics, law, and public administration. What is the relationship between public finance and these four disciplines? In the choice-theoretic approach to public finance, whose chief turn-of-the-century inspiration would be Edgeworth, public finance would be a proper subset of economics. Just as there is a *Journal of Economic Theory,* so there would be a *Journal of Public Economic Theory* to cover that subset of economic theory that dealt with the state. Public finance would look like economic theory, only it would have a specialized subset of subject matter. In this respect, it would be no different from, say, agricultural economics or housing

economics. These are also specialized subsets of economics that are, nonetheless, not anything other than economics. In sharp contrast, a post-cameralist public finance would most surely not be a proper subset of economic theory. Suppose you were to draw a Venn diagram with intersecting circles denoting such fields of study as economics, politics, sociology, public administration, and law. Post-cameralist public finance would cut through all of those fields, and in its own right would be a genuinely multi-disciplinary field of study.

4.   THE ESSAYS TO FOLLOW

The central concern of welfare economics is to explore the relationship between economic welfare and alternative forms for the economic organization of societies. With public finance being the treatment of government within the economic order, the concerns of welfare economics and public finance overlap one another. In Chapter 2, Russell Sobel examines welfare economics in relation to public finance. In particular, Sobel examines the efforts of fiscal scholars to examine the impact of fiscal practices and institutions on efficiency and equity. In the course of this examination, Sobel explores theories of public goods and externality, and considers the impact of government on the distribution of income and wealth.

In their treatment of "Fiscal Constitutionalism" in Chapter 3, Geoffrey Brennan and Giuseppe Eusepi develop a contrast between the constitutional approach to public finance and the orthodox approach. In large measure, this distinction accords pretty well with the distinction between the older and the newer styles of public finance noted above. The orthodox approach to public finance treats fiscal outcomes as reflecting the choices of some benevolent despot. The constitutional approach treats fiscal outcomes as phenomena that emerge out of interaction among a plethora of fiscal participants. The institutional framework within which fiscal outcomes emerge occupies center stage in the constitutional approach. Brennan and Euseppi contrast these two approaches to public finance across a number of fiscal institutions and practices.

Between the start of the twentieth century and the end, the share of national economic activity that flowed through government increased several times over. Where government's typically claimed less than ten percent of GDP at the start of the century, they generally occupied between 30 and 50 percent by the end of the century. In Chapter 4, Thomas Borcherding, Stephen Ferris, and Andrea Garzoni examine the growth of government since 1970. They find that the growth of government has not been so uniformly positive since 1970 as it was over the previous 70 years, and they account for this pattern of growth in terms of both economic variables that operate mainly on the demand side for government services and political variables that operate more on the supply side.

In their exploration of "Rules, Politics, and the Normative Analysis of Taxation" in Chapter 5, Walter Hettich and Stanley Winer contrast different approaches to the development of norms for taxation, and of the treatment of politics within the context of those norms. They distinguish between outcome-oriented rules and process-oriented rules. This distinction maps fairly well into the distinction that Brennan and Eusepi make between the orthodox approach and the constitutional approach in Chapter 3. Outcome-oriented norms are articulated outside of considerations pertaining to political processes, and represent statements about the character of the choices that might be made by some benevolent despot. Process-oriented norms recognize that taxes emerge out of political processes that are suffused with rivalry among interest groups, and seek to constrain the outcomes of those processes by eliminating particularly undesirable outcomes.

Tax revenues are used either to finance programs of public production or to provide transfer payments to recipients. This distinction between public production and transfer payments corresponds to the distinction between the allocative and the distributive branches in Richard's Musgrave (1959) presentation of the *Theory of Public Finance.* In Chapter 6, Randall Holcombe explores "Taxation, Production, and Redistribution" in terms of contrasting the implications of alternative approaches to public finance. Substantial differences result from replacing the presumption that fiscal outcomes are chosen by a benevolent despot with the presumption that those outcomes emerge from within a democratic political process. One famous illustration of benevolent despotism is the Ramsay model of excise taxation, where tax rates vary inversely with the elasticity of demand. Holcombe explains how a rule of uniform tax rates would economize on the various rent-seeking and rent-extraction activities that would exist under a regime of differential taxation, and which would be precluded by a requirement of uniform taxation. In this and in numerous other illustrations, standard propositions about public finance that have been developed within a political context of benevolent despotism are reversed in a democratic setting.

In Chapter 7, Fred Foldvary describes the generation of "Public Revenue from Land Rent." While it is easy to think that this would be just another form of taxation, Foldvary explains why this thought would be mistaken. The value of any piece of land depends on the value of the civic works that lie in proximity to that land. Such things as roads, parks, and schools increase the value of nearby land. Foldvary reports that a bridge across the Hudson River in New York increased adjacent land values by six times the cost of the bridge. The use of changes in land value as a means of financing government could be part of a program by which government participates within the economic process, as against intervening into that process through taxation.

Chapter 8, by Richard Wagner, examines "Debt, Money, and Public Finance." In principle, borrowing and creating money are distinct methods of public finance. They are confounded, however, within modern institutional arrangements. A government whose bonds are bought by citizens is borrowing, but a government whose bonds are bought by the central bank is creating money. While it is linguistically common to speak of governments as being indebted, this common usage is misleading for democratic states where government is simply an intermediary. Public borrowing within a democracy is a means by which state-based intermediation replaces market-based intermediation. This replacement might be universally beneficial or it might be beneficial to some but not to others, with the outcome depending on the institutional arrangements within which political and fiscal outcomes emerge.

The theory of corrective taxation treats instances where taxation seems to serve more a means of regulation than a source of revenue. In Chapter 9, Andy Barnett and Bruce Yandle examine "Regulation by Taxation." They note that the regulatory impact of such taxation often generates differential impacts among interest groups, in addition to generating revenue for governments. Indeed, they claim that such taxation fits the scenario of the Baptists and bootleggers, whereby both support restrictions on liquor sales, though for strongly different reasons. They illustrate their point empirically with reference to liquor taxation within the United States and environmental taxation within the member nations of the Organization for Economic Cooperation and Development.

In Chapter 10 on "Taxation, Black Markets, and other Unintended Consequences," Carla Marchese explores some of the many steps that exist between the establishment of a tax and the actual collection of revenue from that tax. The higher the rate of tax, the stronger will be the incentive to avoid or evade the tax. These efforts can take many particular forms, some legal and some not. This chapter explores tax enforcement and compliance from a variety of angles and perspectives.

In Chapter 11, Scott Hinds, Nicolas Sanchez, and David Schap explore "Public Enterprise: Retrospective Review and Prospective Theory." Prior to the development of public choice theorizing, the bulk of the literature on public enterprise was normative in character. That literature sought to set forth rules govern such things as when public enterprises should be created and what the pricing policies of those enterprises should look like. As public choice theorizing attained prominence, scholars began to realize that the actual conduct of public enterprises often bore little resemblance to the normative formulations. The authors describe this shift from a normative- to a positive-dominated agenda, and look forward to further developments of the positive-dominated agenda.

What is a public enterprise at one time might become a private enterprise later. Or it might have been a private enterprise before it was transformed into

a public enterprise. In Chapter 12, William Peirce examines "Privatization, Nationalization, and Aspects of Transition." There are many possible reasons why private enterprises might be nationalized, or why public enterprises might be privatized. At base, Peirce explains, the shift in organizational form should be explainable in terms of its ability to confer benefits on significant subsets of the population. For instance, a state might privatize to generate a temporary increase in revenue. Alternatively, a state might nationalize to maintain employment among the membership of an influential labor union.

In his wide ranging review of "Social Insurance" in Chapter 13, Heinz Grossekettler examines explanations for the development and growth of social insurance, describes the vast differences across nations, and considers future problems and projections. Social insurance is characterized as having been crafted out of a tension between two opposed sentiments. One is a desire for individual separateness or property. The other is a desire for peace or solidarity within some particular territory. The resolution of these opposing sentiments, moreover, differs greatly throughout the world. This resolution, furthermore, is not of the sort that brings to mind notions of stable equilibrium but rather brings to mind future turbulence.

The term "welfare state" is commonly used to cover both the social insurance programs that Grossekettler examines and programs of poor relief. To be sure, the distinction between social insurance and poor relief is muddied in practice. The metaphor of a safety net, for instance, has been used to characterize both social security types of programs and programs of poor relief. In Chapter 14, Richard Wagner explores "Redistribution, Poor Relief, and the Welfare State." This essay begins with an examination of justifications that have been advanced for the state provision of poor relief, where the alternative to state provision is provision through privately organized charities. Arguments about justification are the province of fiscal philosophers, but the actual creation and operation of programs is the province of political realists. The remainder of this essay examines the problem of state competence in light of the possible clash between the fiscal philosophers and the political realists.

It would surely be hard to support inefficient public programs over efficient programs, at least without mocking linguistic meanings and conventions. Many economists have sought to bring economic principles to bear on the appraisal of the efficiency of public programs. In Chapter 15, Allan Schmid explores a wide variety of issues regarding "Economic Analysis and Efficiency in Public Expenditure." Schmid notes at the outset that efficiency is not simply a matter of technique and calculation; computations of efficiency cannot be constructed that would serve unambiguously as judges of the content of political choices and actions. It is easy enough formally to say that one program is more efficient than another if it generates more output value for the same amount of input. But typically it is not easy to value output, or even to measure it. How,

for instance, is a metric for the amount of education produced within public schools constructed? As Schmid notes, cost-benefit analysis may help to organize our thinking about public programs, but it cannot make our choices for us.

In Chapter 16 on "Local Public Finance," Charles Blankart and Rainald Borck note that while local public finance can be traced back to the middle ages, the field began in earnest only with the publication in 1956 of a paper by Charles Tiebout. Most of the subsequent literature has been concerned in one fashion or another to look for parallels between a network of local governments to order collective activity and competitive markets to order private activity. There are several strands to the literature, all of which Blankart and Borck explore. On the one hand, competitive local governments are able to accommodate more fully differences in preferences across people than would be possible within a single, monopoly government. Competitive governments can also promote experimentation and the generation of knowledge through their ability to serve as types of laboratory experiments of different approaches and policies. On the other hand, a multiplicity of local governments will typically generate externalities, which in turn will create problems of coordination, as well as prompting institutional innovations to deal with those problems.

Competition among local governments is a form of horizontal competition, with those governments competing to attract residents. Within federal systems of government, there is also vertical competition among governments, which arises from the simultaneous citizenship of people in multiple governments. Jean-Michel Josselin and Alain Marciano examine in Chapter 17 a plethora of issues concerning "Federalism and Subsidiarity, in National and International Contexts." The principle of subsidiarity supports the delegation of tasks to the lowest level at which they can be discharged responsibly and effectively. At first glance, subsidiarity might seem to be a principle that operates in favor of lower over higher levels of government within a federation. The practice of federalism, however, is not so neat, and in their examination of why this is so, Josselin and Marciano distinguish between federating a nation (as illustrated by the United States) and federating a set of nations (as illustrated by the European Union).

In Chapter 18 on "Fiscal Sociology: What For?" Jürgen Backhaus locates fiscal sociology as the contemporary continuation of the tradition of Staatswissenschaften. Economics and politics are thoroughly co-mingled in contemporary societies, and the disciplinary separation of analytical spheres calls also for efforts at unification. Backhaus explains how fiscal sociology can serve as the contemporary continuation of Staatswissenschaften, by promoting an integrative treatment of the economic, political, legal, and sociological aspects of public finance. He does this by generating an entire alphabet of concepts that fiscal sociology can illuminate.

REFERENCES

Backhaus, Jürgen G. (1993). "The German Economic Tradition: From Cameralism to the Verein für Socialpolitik." In *Political Economy and National Realities,* Albertone, Manuela and Masoero, Alberto, eds., pp. 329-356. Torino: Fondazione Luigi Einaudi, 1994.

Backhaus, Jürgen G. (1992). "The State as a Club: A Perspective for Public Finance in a Prosperous Democracy." *Journal of Public Finance and Public Choice* 10: 3-16.

Backhaus, Jürgen G. (2002). "Old or New Public Finance? A Plea for the Tried and True." *Public Finance Review* 30: 612-45.

Backhaus, Jürgen G. and Wagner, Richard E. (1987). "The Cameralists: A Public Choice Perspective. *Public Choice* 53: 3-20.

Brennan, Geoffrey and Buchanan, James M. (1980). *The Power to Tax: Analytical Foundations of a Fiscal Constitution.* Cambridge: Cambridge University Press.

Buchanan, J. M. (1960). "The Italian Tradition in Fiscal Theory." pp. 24-74 of his *Fiscal Theory and Political Economy.* Chapel Hill: University of North Carolina Press.

De Viti de Marco, Antonio. (1936). *First Principles of Public Finance.* London: Jonathan Cape.

Edgeworth, F. Y. (1897). "The Pure Theory of Taxation." In *Classics in the Theory of Public Finance,* Musgrave, R. A. and Peacock, A. T., eds., pp. 119-136. London: Macmillan, 1958.

Foldvary, Fred. (1994). *Public Goods and Private Communities.* Hants, UK: Edward Elgar.

Goode, Richard. (1970). "Public Finance in *the International Encyclopedia of the Social Sciences:* A Review Article." *Journal of Economic Literature* 8: 27-34.

Humpert, Magdalene. (1937). *Bibliographie der Kameralwissenschaften.* Köln: Karl Schroeder.

Justi, Johann. Heinrich Gottlob von. (1771). *Natur und Wesen der Staaten.* Darmstadt: Scientia Verlag Aalen, 1969. [Reprint edition.]

MacCallum, Spencer H. (1970). *The Art of Community.* Menlo Park, CA: Institute for Humane Studies.

Mirrlees, J. A. (1994). "Optimal Taxation and Government Finance." In *Modern Public Finance,* Quigley, J. M. and Smolensky, E., eds., pp. 213-231. Cambridge: Harvard University Press.

Musgrave, R. A. (1959). *The Theory of Public Finance.* New York: McGraw-Hill.

Ramsey, F. P. (1927). "A Contribution to the Theory of Taxation." *Economic Journal* 37: 47-61.

Schumpeter, Joseph A. (1954). *History of Economic Analysis.* New York: Oxford University Press.

Small, Albion. (1909). *The Cameralists: The Pioneers of German Social Polity.* New York: Burt Franklin.

Smith, Adam. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations.* New York: Modern Library, 1937.

Tribe, Keith. (1984). "Cameralism and the Science of Government." *Journal of Modern History* 56: 263-84.

Wagner, Adolf. (1883). *Finanzwissenschaft,* 3rd ed. Leipzig, C. F. Winter.

Wagner, Richard E. (1997). "Choice, Exchange, and Public Finance." *American Economic Review,* Proceedings 87: 160-63.

Wagner, Richard E. (2002). "Some Institutional Problematics of Excess Burden Analytics." *Public Finance Review* 30: 531-45.

Wicksell, Knut. (1896). "A New Principle of Just Taxation." In *Classics in the Theory of Public Finance,* Musgrave, R. A. and Peacock, A. T., eds., pp. 72-118. London: Macmillan, 1958.

Chapter 2

WELFARE ECONOMICS AND PUBLIC FINANCE

Russell S. Sobel
West Virginia University
rsobel2@wvu.edu

Abstract	This contribution deals firstly with the differences between market action and government action, and then explores the justification for government intervention based on concepts of economic efficiency and equity. The chapter then proceeds to discuss individual cases in which unregulated private market outcomes are generally considered to violate this criterion.
Keywords:	Equity, economic efficiency, economic stabilization, market failure, monetary stability, welfare economics
JEL classification:	D60, H11

1.   INTRODUCTION

In a market economy, it is commonly accepted that the role of government should be limited. This philosophical approach not only dominates economic thinking back to the time of Adam Smith's *Wealth of Nations* in 1776, but can also be clearly seen in eighteenth-century political philosophy in the writings of Locke, Jefferson, and Madison, among others. It is a philosophical approach that is plainly expressed in the U.S. Constitution adopted in 1789.[1] The modern interpretation of the principle of limited government within the field of economics envisions a more active role for government than the founding fathers would have held. It is, however, still based in the idea that public sector intervention should be limited. In particular, government intervention should be limited to cases in which the outcome of the private unregulated market is somehow judged to be undesirable. That is, in each case, the market outcome is compared to some ideal and only when it fails to meet that ideal is there a role for government intervention.

In modern economic analysis, the two criteria generally used to judge a market outcome are efficiency and equity. Efficiency is defined as economic (or Pareto) efficiency, while equity deals with the more ambiguous issue of fairness. These two criteria differ substantially as the first (efficiency) is a positive, objective criterion, while the other (equity) is a normative, subjective criterion. Because of this difference, arguments for government intervention in cases when markets fail to achieve efficiency are somewhat less controversial than are arguments for government intervention based on equity considerations. It is worth explicitly noting that the commonly used term "market failure" corresponds *only* to cases in which the private unregulated market outcome fails to meet the conditions for economic efficiency and is not generally used for judgments on equity grounds.[2]

Economic thinking about the role of government in the economy has undergone a drastic change over the past three decades due primarily to the insights provided by public choice analysis. It was once thought that any case in which a market failed to meet the conditions for economic efficiency *necessarily* implied that the government should intervene and move the market toward the efficient outcome. Recent economic thinking incorporates the idea that public sector institutions are also imperfect, that there is a cost of using them, and thus there is no *a priori* reason to believe that government intervention into an imperfect market will necessarily lead to a more efficient outcome. This is perhaps best illustrated in the following quote from George Stigler:

> A famous theorem in economics states that a competitive enterprise economy will produce the largest possible income from a given stock of resources. No real economy meets the exact conditions of the theorem, and all real economies will fall short of the ideal economy— a difference called "market failure." In my view, however, the degree of "market failure" for the American economy is much smaller than the "political failure" arising from the imperfections of economic policies found in real political systems. The merits of laissez-faire rest less upon its famous theoretical foundations than upon its advantages over the actual performance of rival forms of economic organization.[3]

Indeed, it is now accepted that in some cases an unregulated "bad" market outcome may still be preferable to the one achieved with government intervention.[4] The burden has shifted from one in which government involvement was justified in all cases of imperfect market outcomes to one in which government involvement is justified only in cases where the potentially imperfect outcome with government involvement is likely to be better than the imperfect outcome with an unregulated private market. Thus, modern public sector economists tend to be in favor of an even more limited role of government than were public sector economists of the past.

This chapter proceeds by first discussing the differences between market action and government action, and then exploring the justification for government intervention based on concepts of economic efficiency and equity. The

chapter then proceeds to discuss individual cases in which unregulated private market outcomes are generally considered to violate these criterion.

2. THE DIFFERENCE BETWEEN MARKET ACTION AND GOVERNMENT ACTION

The private sector (markets) and the public sector (government) may simply be thought of as two alternative institutions that can be used to allocate scarce resources in an economy. In a *market* economy, characterized by private ownership, it is important to remember that these resources are not owned collectively by society, but rather are owned privately by individuals. The market process that allocates these resources works through the voluntary, uncoerced specialization and exchange undertaken by individual owners. In contrast, collective action undertaken through the public sector uses the coercive powers of government to alter the choices of individual owners. This is the first of two fundamental differences between market action and government action—the reliance on voluntary choice versus coercion to allocate resources. When market exchange occurs it is clear that both parties have been made better off (or were both expecting to be made better off), while with government action it is frequently the case that some parties have been made better off while others have been made worse off.[5]

The second fundamental difference between market action and government action rests in the nature of planning and choice. In the public sector planning is done centrally, while in private markets planning is done individually. Government intervention can thus be thought of as replacing individual planning with central planning. In markets, individuals are left to make choices based on the personal costs and benefits they face according to their individual preferences. When action is done through the public sector, the choices and decisions must be made collectively. Collective choice is a much more difficult process than individual choice as it requires a mechanism for aggregating the preferences of many diverse individuals. To make good collective choices requires registering or knowing a vast amount of information about individual preferences. The fact that no single central planner could possibly know all the information necessary to make these good choices was a key element of F.A. Hayek's (1945) defense of capitalism over socialism. In modern market based economies, democratic voting procedures, rather than the selection of a knowledgeable central planner, is generally used as the process to make collective choices. These voting rules, however, inherently have problems with registering the intensity of preferences, getting individuals to truthfully reveal their preferences, and providing enough incentive for voters to become well informed about the choices they must make.[6]

Models of public sector intervention in cases of market failure have historically modeled government as being represented by a socially benevolent dictator who had all the information necessary to make changes that would improve the efficiency of resource allocation. Modern day economic analysis, however, generally models the process of collective choice as one dominated by rationally ignorant voters, powerful special interest groups, vote-maximizing elected officials, and budget-maximizing bureaucrats. It should be apparent that this has important implications for government intervention, both to correct market failure and to achieve normative equity goals. Interest groups and bureaucrats will tend to cloak their self-interested demands for transfers, budgets, and legislation as policies to address market failures or equity goals, even when that is not the true intention of the policy. For this reason, stringent constraints on government intervention and regulation appear necessary.

3.   THE CONCEPT OF ECONOMIC EFFICIENCY

Within the neoclassical economic paradigm, economic efficiency is the benchmark by which both market outcomes and government intervention are judged. Economic efficiency requires two conditions be met:

(1) all actions generating more social benefits than costs should be undertaken, and
(2) no actions generating more social costs than benefits should be undertaken.

If both of these conditions are met, a Pareto Optimal allocation will be attained—that is, one in which it is impossible to reallocate resources in such a way to make at least one person better off without harming another person.[7]

When market exchange occurs it is clear that both parties have been made better off, while when government action occurs it is frequently the case that some parties have been made better off while others have been made worse off. If all parties to an exchange benefit it is clear that the action is consistent with efficiency. In cases where government intervention benefits some parties and harms others, the efficiency implications are not so obvious. The traditional metric by which such actions are judged is the "potential Pareto criterion" (sometimes referred to as the Hicks-Kaldor criterion).[8] The potential Pareto criterion is met if enough benefits are generated such that it would be hypothetically possible for the winners to completely compensate the losers. In essence, the potential Pareto criterion amounts to a cost/benefit test for government intervention. It is important to note that substantial issues arise with a strict application of this rule. For example, if the benefits of building a road exceed the losses to property owners from taking their property for use in construction, the potential Pareto criterion would justify taking the property for

public use *regardless* of whether any compensation was paid to the owners at all.

Almost exclusively in public finance, the efficiency criterion is applied to whether the quantity of some economic activity is the efficient quantity, and the benchmark efficient quantity is generally derived or illustrated in a supply and demand diagram in which the supply curve measures the marginal social cost of the activity, while the demand curve measures the marginal social benefit of that activity. This is illustrated in Figure 1 where MB_S and MC_S are the marginal social benefit and marginal social cost respectively.

In Figure 1, Q* corresponds to the efficient output level. All units up to Q* satisfy condition (1) listed above because they all generate more social benefits than costs. Units beyond Q* should not be produced given condition (2) listed above because they generate less social benefits than costs.

Private individuals acting in markets make decisions to buy and sell based on the private (or personal) costs and benefits they face. If all of the costs and benefits from an activity are isolated to only the parties privately involved in the transaction, then it will be the case that the private costs and benefits on which the market decision is based fully reflect all of the social costs and benefits of the action.[9] More precisely, actual market outcomes are determined by the intersection of demand and supply curves that reflect only the marginal private benefits (MB_P) and marginal private costs (MC_P) of the activity. Thus, in cases where the marginal private benefits equal the marginal social benefits $(MB_P = MB_S)$ and the marginal private costs equal the marginal social costs $(MC_P = MC_S)$, the equilibrium quantity produced in a competitive private market will be precisely the Q* shown in Figure 1. Cases in which private and social costs (or benefits) diverge will result in a private market outcome that is not consistent with the efficient level of output. These are cases of market failure that are to be explored in further detail in this chapter.
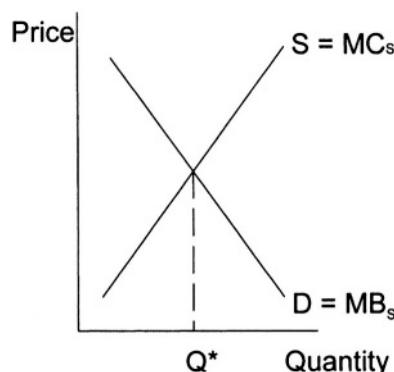

FIGURE 1. Market Efficiency.

4.   THE CONCEPT OF EQUITY

The intervention of government into markets to address equity concerns is a more controversial issue than is the intervention of government into markets to correct cases of market failure to achieve economic efficiency. At the heart of this controversy is the lack of a positive, objective definition of equity. Even the best attempts in the economics literature to objectively define fairness have failed upon closer scrutiny.[10]

Many modern scholars argue that the fairness of an outcome cannot be determined without knowledge of the process that determined the outcome.[11] More precisely, they adopt a procedural theory of fairness in which a fair outcome is defined as one that is the result of a fair process. Within this framework it is possible to have outcomes that are clearly unequal, but are fair nonetheless because they were the result of a fair process. Correspondingly, it is possible to have apparently equal outcomes that are unfair because they are the result of an unfair process. If one perceives the market as a fair process, then any distribution of income or wealth that results from it must, by definition, be fair.

The immense difficulty in applying these different fairness concepts can be seen when analyzing the merits of alternative tax proposals. If one views taxes as a way of allocating the cost of financing government across individuals, it appears fair to assign taxes in accordance with ability to pay (although the degree to which taxes rise with ability to pay would still be an issue). Alternatively, if one views taxes as the price citizens pay for government output, then taxation according to benefits received appears to be the fair method of assigning taxes. A citizen who does not benefit from a particular government program should not be forced to pay for it, regardless of their income. This example makes it clear that even in the restricted area of tax policy, the concepts of fairness and equity are difficult to define in a manner that is considered agreeable by everyone.

While there is clearly popular support for democratic governments to intervene into markets for equity reasons, it is less clear whose definition of equity should be used as the basis. In cases where government involvement to achieve equity goals detracts from the efficiency of markets, the equity justification may stand at odds with the logic of using government to promote economic efficiency. Perhaps ironically, one could apparently argue on this same ground that there could be equity based justifications for *not* allowing government to correct a market failure if reaching the efficient outcome would detract from a stated equity goal.

At the heart of social welfare analysis is the idea that while there are many possible efficient allocations of resources (imagine all the points along the contract curve in an Edgeworth box for example), not all of these points are

equally preferred from the standpoint of equity. Operationally, it has been standard practice in public finance for economists to incorporate equity goals into economic models through an explicit representation of a social welfare function, the social welfare function simply being some algebraic transformation of the utility levels of the members of the society.[12] The social welfare function can then be maximized subject to the production or other constraints imposed on the economy to obtain the solutions that maximize social welfare. Because these models require arbitrary weights to be placed on the utility levels of different members of society, the value of such mathematical exercises depends on whether one agrees or not with the subjective weighting choices made by the author of the model.

Regardless of whether the justification is on efficiency or equity grounds, there are several widely accepted areas in which government intervention might be justified. The remainder of this chapter is devoted to more in depth discussions of these areas, which are (1) public goods, (2) externalities, (3) monopoly, (4) incomplete information, (5) economic stabilization, and (6) redistribution.

5.   PUBLIC GOODS

The first case of market failure that will be discussed is that of public goods. A public good is defined as any good that is both joint-in-consumption (sometimes called nonrival-in-consumption) and nonexcludable. To be a public good, a good must have *both* of these characteristics. This section first defines these two characteristics and explores the conditions necessary for the efficient provision of a public good. It then proceeds to explain why a private market may fail to efficiently produce a public good and whether a real world public sector is capable of producing the efficient level of a public good.

A good is "joint-in-consumption" if the consumption of the good by one individual does not lower the amount of the good available to others. Each unit of a good that is joint-in-consumption can be shared by all consumers, and the marginal cost of providing the good to one additional user is precisely zero. A radio broadcast signal provides an example of a good that is joint-in-consumption. If an additional listener turns on their radio, they may receive the signal without detracting from the amount of the signal available to others. In addition, if the population within the listening area were to increase, the broadcast signal is available to these additional listeners with no additional cost of production. A public good that is subject to congestion as the number of users grows is sometimes referred to as an impure public good. A road in a rural area may have so little traffic that the addition of one additional car does not detract from the availability, or value, of the road to other users—so it is joint-in-consumption. That same road placed in the downtown of a

metropolitan area, however, may become congested and lose its jointness-in-consumption—and thus would no longer exhibit this characteristic. Thus, the publicness of a good depends not only on the good itself, but also on the environment in which it is consumed. Thus, an impure public good may be a public good in some situations, but not in others.

A good is nonexcludable if it is impossible (or at least prohibitively costly) to exclude nonpaying consumers from receiving the good. Consider, for example, a Fourth of July fireworks display provided in a public park. If an admission fee to the park were charged, some individuals might choose to watch the fireworks display from just outside the park to avoid paying the entry fee. In cases where individuals may still receive the benefit from the good without paying, they will have an incentive to do so, particularly in cases where the lack of their individual payment does not have a significant impact on the total quantity of the good provided. This potential for "free riding" by users of the good is the source of the potential market failure in the case of a public good. If a private firm cannot exclude nonpaying customers, their revenue will not fully reflect the social benefit derived from the production of the good. This will be a case in which the marginal private benefit of the activity reflected in the market demand curve is less than the marginal social benefit of the activity ($MB_P < MB_S$). Because free riding lowers the private benefit to the firm of producing the good, it will be supplied in a less than optimal quantity—if it is supplied at all. This free-rider problem is at the heart of the arguments for market failure, and public sector provision, in the case of public goods. We will return to the issue of whether markets can, in some cases, find ways to overcome the free-rider problem and thus efficiently produce public goods after first deriving the necessary conditions for the efficient provision of a public good.

As is the case with any good, the efficient level of production may be found by equating the marginal social benefit and marginal social cost curves as was done in Figure 1. There is one fundamental difference, however, in the construction of the marginal social benefit curve between private and public goods. To construct a marginal social benefit curve (the market demand curve) in the case of a pure private good, it is necessary to *horizontally* sum all the marginal benefit curves of the individuals in the market (the individual demand curves). In the case of a good that is joint-in-consumption (regardless of its excludability), it is necessary to *vertically* sum all the marginal benefit curves of the individuals in the market. The reason for the difference is that in the case of a good that is joint-in-consumption each unit is jointly shared by all and thus the total social benefit produced from a given unit is the sum of the benefits derived by all individual consumers who share in the consumption of that unit. In the case of a private good, each unit is rival-in-consumption so that the total social benefit produced by the good is only the private benefit received by the

single individual who obtains and consumes the good. The construction of the market demand curve (D) or marginal social benefit curve (MB_S) for a public good in a society of three individuals, Larry, Moe, and Curly with individual demand curves given by d_L, d_M, and d_C, is shown in Figure 2.

The economically efficient quantity of this public good (Q*) is illustrated in Figure 2 and it is found by the intersection of the marginal social benefit curve with the marginal social cost curve. For simplicity, here it is assumed that there is constant marginal social cost in the provision of the public good. The condition that must be present for the efficient provision of a public good is that the sum of the marginal rates of substitution across all individuals (here equivalently modeled as the individual marginal benefits) must be equal to the marginal cost of production (or equivalently, the marginal rate of transformation in a general equilibrium model). This condition is often referred to as the "Samuelsonian condition" for the efficient provision of a pubic good because Samuelson (1954) was the first to formally derive it. His original article was followed by the publication of the diagrammatic representation of this condition in Samuelson (1955).[13]

How will the output level of this public good in a private unregulated market compare to the efficient quantity shown in Figure 2? Because of the nonexcludability of the good, the free-rider problem discussed above will result in a private provision equilibrium in which the quantity produced is less than the efficient quantity.[14] Despite the rather clear implications of the neoclassical maximization model for the inefficiency of private provision of public goods, many scholars are very critical of the real-world applicability of this model.
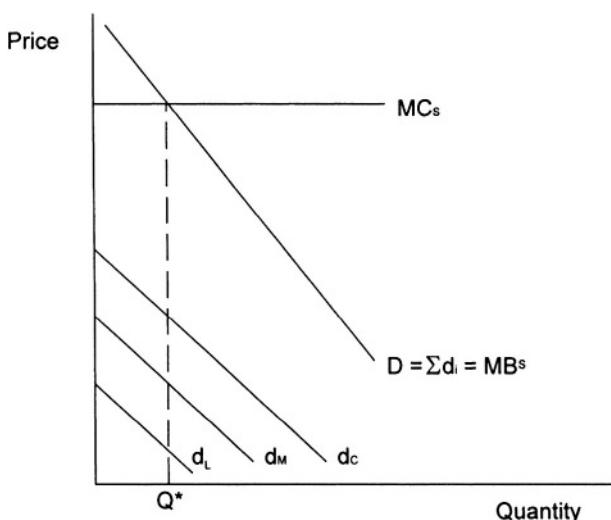

FIGURE 2. Efficient Provision of a Public Good.

Radio broadcasts, for example, meet both conditions for a public good, but rather than the private market underproviding broadcasts, the Federal Communications Commission actually restricts the number of radio stations allowed in the private market under the logic that the private market would otherwise *oversupply* radio broadcasts. Similarly, lighthouses were traditionally listed as a textbook case of a public good. Subsequent research by Coase (1974) and Peacock (1979), however, has found that lighthouses in nineteenth-century England were indeed privately provided. Finally, Holcombe (1996) points out that Bill Gates became the richest man in the world producing a good, computer software, that can be argued to meet the conditions for being a public good (particularly prior to the development of copy protection technology). While examples such as these don't prove that markets can provide public goods efficiently, they certainly cast doubt on the radical claim that markets can not provide public goods.

When considering whether the private market can efficiently produce public goods, it is important to remember that cases of market failure represent cases in which the full gains from trade have not been realized. Thus, cases of market failure represent profit opportunities for entrepreneurs who can find innovative ways to overcome the sources of the market failure. Because the source of underprovision is the free-rider problem, innovative methods for overcoming this problem can allow private markets to efficiently provide public goods. In the case of radio broadcasts, for example, the use of advertising, rather than direct sale of the broadcast to consumers, allows the industry to overcome the free-rider problem. In the case of lighthouses in nineteenth-century England, rather than funding them by sale of the services directly to ships, the services were sold to nearby ports who found that a lighthouse was essential to be able to attract ships and compete with other ports.

In addition to finding alternative payment mechanisms to circumvent charging the final consumer, another method by which private markets can overcome the free-rider problem is through bundling the public good with another good or service as a tie-in sale.[15] In the case of computer software, for example, the sale of customer support and manuals for the software are bundled with the purchase of the software itself, giving consumers an incentive to pay for the software to receive these other benefits. Shopping malls often provide public goods such as restrooms, common areas with benches and fountains, and security that are not financed by charging individual users. Instead, their provision is financed through the higher lease or rental prices for mall space that results from attracting more customers to the mall. Similarly, neighborhood associations, condominiums, and apartment complexes often provide public goods (such as pools, parks, meeting facilities, fitness facilities, or playgrounds) for their residents that are financed through the higher rental rates (or homeowner

fees) that result from the increased value to residents of having these goods provided for their use.

Examples such as these are used by many authors to question, at a fundamental level, the applicability of the standard neoclassical market failure argument in the case of public goods because it is derived under such restrictive conditions and assumptions about the allowable means of financing the provision of the public good, and because it ignores the great incentive given to private markets to overcome cases of market failure.[16] The ability of private markets to efficiently provide even the most fundamental of public goods, such as a legal system, courts, and contract enforcement, has been shown by Benson (1990). Clearly much additional research is needed to fully understand the conditions under which the private provision of public goods is possible and efficient. In addition, there remains substantial debate as to whether there are many goods that would classify as pure public goods in the first place.[17]

Next, it is worth considering the issue of whether real world public sector institutions are capable of producing the efficient quantity of a public good. While this might not be much of a problem for a benevolent, fully-informed central planner, it can be quite a challenge for a real-world political institution. If collective choices about the provision of public goods are made under majority rule voting, it is possible to derive the amount of the public good that will be supplied by government using the median voter theorem.[18] Let us return to the example of a public good in a community of three individuals that was illustrated in Figure 2. Assume, momentarily, that the good will be financed through a system in which each voter pays one-third of the marginal cost of production. Given this cost sharing agreement, the most preferred quantities of each of the three voters, shown by Q_L, Q_M, and Q_C, are shown in Figure 3.

Under simple majority-rule voting, the median voter theorem applies, so that the median voter's most preferred outcome wins, because it will beat all other alternatives in pair-wise majority voting. Here, the median voter is Moe, so the level of production shown by the quantity Q_M would be produced through the collective choice mechanism. But how will Q_M compare to the efficient level of production of the public good given by Q*? Only in the case where the median voter's tax share exactly equals his or her share of the marginal benefit of production will Q_M equal Q*. If the median voter's tax share is greater than his or her benefit share, Q_M will be less than Q* and if the median voter's tax share is less than his or her benefit share, Q_M will exceed Q*. It is this final case that is illustrated in Figure 3.

The general principle illustrated here is that the closer are tax shares to reflecting the benefits individuals receive from public goods, the closer will be the production of the good by the public sector to the efficient quantity. A tax situation in which each person is charged a tax price equal to their precise marginal benefit at the efficient output level is known as Lindahl pricing, after

the work of Lindahl (1919) that was later formalized by Johansen (1963). The Lindahl tax prices for the three individuals are shown in Figure 4 as T_L, T_M, and T_C.

Note, however, that all that really matters for efficient provision under majority rule voting is whether the *median voter's* tax share equals his or her benefit share. From an efficiency standpoint, whether this is true for the other
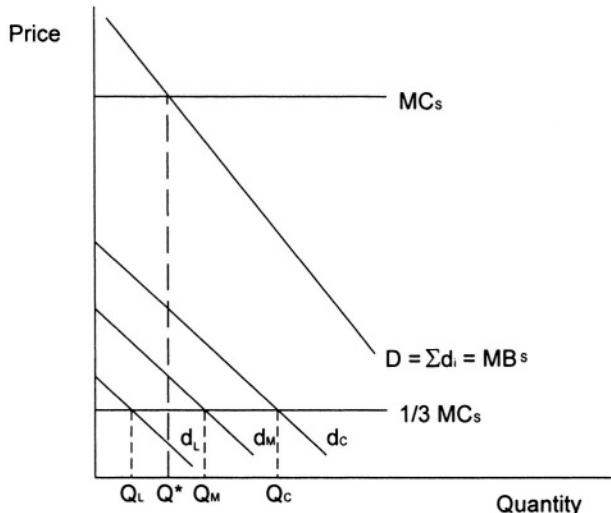
FIGURE 3. Public Sector Provision of a Public Good Under the Median Voter Model.
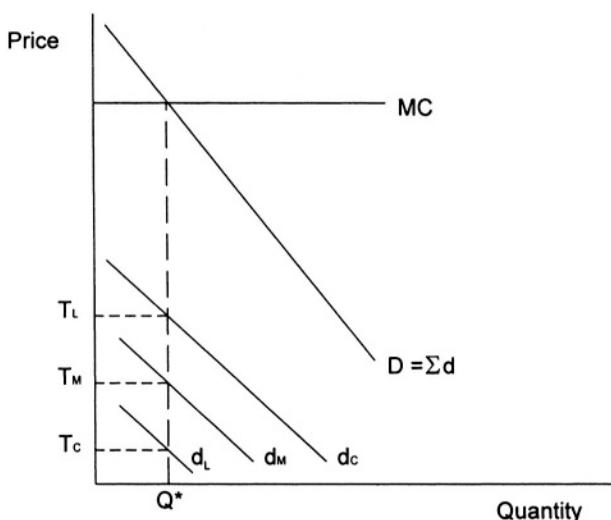
FIGURE 4. Determining Lindahl Tax Prices for a Public Good.

individuals in the society doesn't affect the outcome.[19] Thus, Lindahl pricing
is not a necessary condition for efficient provision, but is rather a sufficient
condition. If this simple majority rule voting model is an accurate representa-
tion of the collective choice process, then the issue as to whether the public
sector can efficiently provide a public good simply depends on the degree to
which the median voter's tax share approximates his or her share of the mar-
ginal benefit of a public good's provision. There are two significant problems
with using this as a guide to tax policy, however. First, it is impossible to ac-
curately estimate the benefit shares of individual citizens, and second, if this
is the announced method for determining tax shares, individuals have a strong
incentive to misrepresent their true preference for the public good in order to
lower their tax burden (by claiming they get less benefits than they really do
from the public good).[20] In reality, when one considers the remote chances
that the median voter's true tax share approximates his or her benefit share, it's
clear that just like in the market sector, the efficient provision of public goods
by government is unlikely. In any particular case the issue is thus whether the
potentially inefficient market outcome is closer or further from efficiency than
the potentially inefficient government outcome.

   Before moving on to the next area of market failure, it is worth considering
one simple extension of the model of public sector provision above. It is now
widely accepted in economics that the public sector bureaus charged with the
actual production of these public goods are far from efficient. In particular, in-
dividuals within these bureaus have very little incentive to control costs. Since
the incentives for internal efficiency are less in public sector bureaus than in
private firms, it is the case that public sector provision of the good will gen-
erally be more costly than private provision of the same good.[21] In addition,
following the work of Niskanen (1968, 1971), the individuals in charge of
public sector bureaus are often modeled as attempting to maximize the size of
their budgets.[22] By presenting "all-or-nothing" type proposals to their sponsor
or funding agency, they can secure a budget that is significantly larger than the
sponsor's most preferred amount. In fact, in a case where the demand curve of
the sponsor is linear, the bureau can obtain funding for a quantity that is up to
twice as large as the sponsor's most preferred quantity. If we briefly consider a
situation in which the median voter's demand curve (d_M) is used to represent
the preferences of the sponsor, and assume the case of Lindahl pricing (that
in the previous analysis resulted in efficient public sector provision in which
the median voter's most preferred quantity Q_M was equal to the efficient quan-
tity Q*), Figure 5 shows the relationship between the quantity preferred by
the median voter (Q_M) and the quantity that would be supplied by a budget
maximizing bureau (Q_B).[23] This is obtained by the construction of an "all or
nothing" demand curve, shown in Figure 5 by d_{AON}.
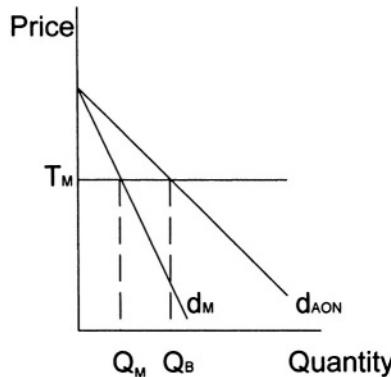
FIGURE 5. Public Sector Provision of a Public Good Under the Bureaucracy Model.

This analysis shows even if tax shares could be allocated in such a way that the efficient quantity of the public good was most preferred by the median voter, potential problems with the incentives of the public sector bureaucracies providing the good may cause the output of the public good to diverge from the efficient quantity. If we were to add into this analysis the fact that the median voter's tax share was probably not equal to the accurate Lindahl tax price, so that Q_M in Figure 5 wasn't the efficient quantity to begin with, it becomes even more clear that efficient public sector provision of public goods is indeed unlikely.[24]

To summarize, this section first defined a public good, proceeded to show the method for determining the efficient provision level for a public good, and then discussed how the free-rider problem created the potential for market failure—in that markets might tend to underproduce public goods. Cases in which markets seem to apparently produce public goods fairly well were discussed, and then cases in which government production was likely to diverge from efficiency were presented. It seems clear that if efficient production is the goal, that simply demonstrating a good meets the criteria for a public good is not sufficient to warrant government intervention. Indeed, there appears to be an additional burden of proof that the government provision is likely to improve upon the private market outcome.

6.   EXTERNALITIES

The second area of market failure to be considered is the case of externalities. Generally an externality may be thought of as a case in which a non-consenting third party is affected, either positively or negatively, by an action undertaken by other individuals. An important distinction, however, arises between cases of pecuniary and technological externalities.[25] A pecuniary externality is a third-party effect that occurs through the pricing system, while

a technological externality is a third-party effect that occurs outside the pricing system. As an example, a McDonald's opening up next door to a Burger King would lower the profits of the existing Burger King. Because this occurs through the market pricing system, this would be considered a pecuniary externality. Because they occur within markets, pecuniary externalities do not create market failures, and are not a justification for government intervention. In fact, the ability of some firms to enter and compete with existing firms (the infliction of these pecuniary externalities) is *necessary* for market efficiency. Holcombe and Sobel (2001) discuss this distinction between pecuniary and technological externalities in more detail and show that when the government intervenes to compensate for pecuniary externalities that it actually moves market outcomes away from efficiency. While the distinction between pecuniary and technological externalities is well developed in the case of the production of business firms, Holcombe and Sobel (2000) provide the first treatment of this difference applied to externalities between individuals. Their analysis suggests that interdependent utility functions are a case of pecuniary externalities that do not require government corrective action.

A technological externality exists only in cases where there is a missing market, an undefined property right, or an unpriced resource at play. Air pollution, water pollution, and overutilization of common property resources are examples. If, to alter the previous example, McDonald's were to emit pollution into the air that interfered with Burger King's ability to produce its hamburgers, this would be a case of a technological externality. Technological externalities may either be positive (external benefits) or negative (external costs). In cases where technological externalities exist, there will be a divergence between the marginal social benefits (or costs) and the marginal private benefits (or costs). Figures 6a and 6b illustrate these two cases.
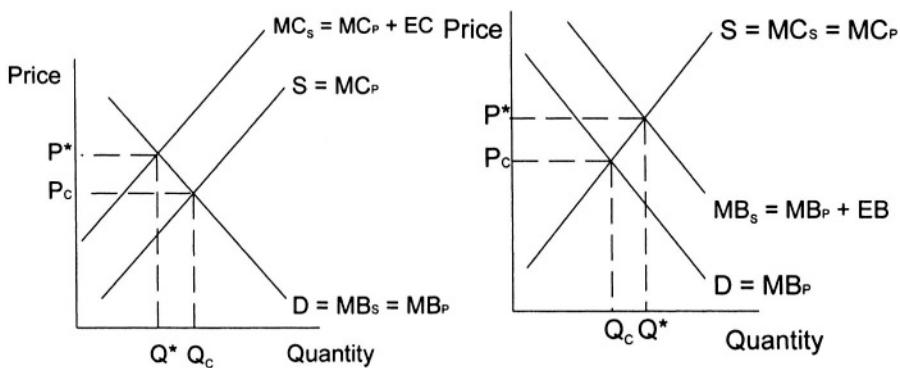

(a) External Cost   (b) External Benefit

FIGURE 6. Private Market Failure in the Case of Technological Externalities.

Figure 6a illustrates the standard neoclassical analysis in the case of an external cost in the amount of EC per unit of the good produced (assuming no externalities on the benefit side of the market). The total marginal social cost (MC_S) is equal to the marginal private cost (MC_P) plus the external cost (EC). Because the private market responds only to the private costs involved, which are an understatement of the true social costs, the competitive private market outcome Q_C will be greater than the efficient outcome of Q*. The good will also be underpriced (that is, $P_C < P^*$) because the market does not fully incorporate the true social cost of production. Figure 6b illustrates the analogous case for external benefits in the amount EB per unit of the good produced (assuming no externalities on the cost side of the market). Here total marginal social benefit (MB_S) is equal to the marginal private benefit (MB_P) plus the external benefit (EB). Because the private market responds only to the private benefits involved, which are an understatement of the true social benefits, the competitive private market outcome Q_C will be less than the efficient outcome of Q*. As in the case of external costs, a good with external benefits will also be underpriced ($P_C < P^*$).

The above analysis was for the case in which the external cost (or benefit) was a constant amount for each unit of the good produced (thus the parallel, vertical shifts in the supply and demand curves). When this condition is violated, it is possible that the competitive market outcome may still be efficient in the presence of an externality. In particular, consider a case in which the external cost is, say $5 on the first unit produced, $4 on the second unit, $3 on the third unit, and so forth until the marginal external cost goes to $0 on the sixth and subsequent units. If water pollution from a firm is killing fish in a lake, for example, it is likely the case that after a certain level of production that additional units produced (and additional pollution emitted) do not create any additional marginal damage. In the case of external benefits, say for example an individual's choice of educational attainment, it may be the case that the external benefits generated by the first few years of schooling are large, but that as additional years of schooling are acquired, these external benefits eventually go to zero beyond some educational level. If the marginal external costs or benefits fall to zero *before* the level of production that would be provided by a competitive private market, then there will be no relevant externality at the margin, and thus no market failure. This case of "inframarginal externalities" is illustrated in Figures 7a and 7b.

As is illustrated in Figures 7a and 7b, when the externalities are inframarginal, the private market outcome is efficient because the externality is not relevant at the margin (i.e., at the equilibrium quantity). To distinguish the case in which there is an externality relevant at the margin, such as in the cases shown in Figures 6a and 6b, those are sometimes referred to as "Pareto-relevant externalities," to contrast them with the case of inframarginal externalities.[26] It
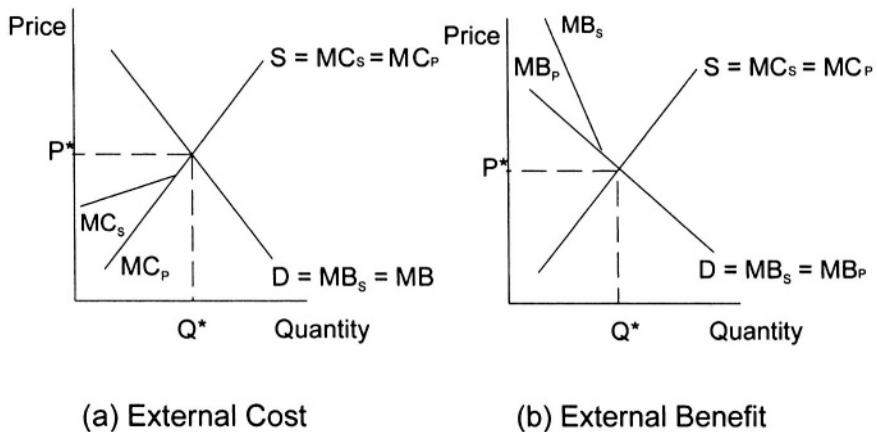
FIGURE 7. Private Market Efficiency in the Case of Inframarginal Externalities.

is possible, however, that in the case of an inframarginal externality if demand (or supply) were to decline, that the externality would become Pareto relevant.

Based on the seminal work of A.C. Pigou, for decades the dominant thought was that in the case of a Pareto-relevant externality, that government intervention in the form of a tax or subsidy would be required to move the market toward the efficient outcome.[27] Returning to the case of an external cost in Figure 6a, the government could impose a per-unit tax in the amount of the external cost. The private market supply curve would shift up vertically by the amount of the per-unit tax, and as long as the per-unit tax (T) was equal to the amount of the external cost (EC) created per unit, then the new private market supply curve would mirror the true marginal social cost curve, and the market equilibrium quantity would move to the efficient output level of Q*.[28] In the case of an external benefit as in Figure 6b, the government could grant a per unit subsidy (S) equal to the amount of the external benefit (EB) created by each unit. This would shift the private market demand curve upward vertically by the amount of the subsidy. The resulting demand curve inclusive of the subsidy would mirror the true marginal social benefit curve, and the market equilibrium quantity would move to the efficient output level of Q*.[29]

The imposition of these "Pigovian" taxes and subsidies in practice is difficult, however. Proper policy requires that the government officials in charge of determining the tax and subsidy amounts have knowledge of the exact amounts of the true external costs or external benefits in the market. Furthermore, individuals would have an incentive to misrepresent their true preferences in this case if the information they were required to provide to the government impacts their tax or subsidy amount. Finally, even if it were possible to know the true external costs or benefits, one must ask what incentive government would have to impose taxes or subsidies in those amounts. If the government were

allowed to tax (or subsidize) a particular market, the tax (or subsidy) imposed would likely reflect many political factors other than the externality. A real-world legislature might, for example, impose the revenue maximizing per-unit tax, or increase the subsidy beyond the amount of the external benefit in an attempt to win votes for an upcoming reelection. Additionally, because a firm or individual would lobby just as hard to avoid or prevent a $100 technological externality as a $100 pecuniary externality, a vote-seeking politician may attempt to enact policies that prevent or compensate for both types of externalities, and as Holcombe and Sobel (2001) show, government intervention to prevent or correct pecuniary externalities results in less, rather than more, efficient market outcomes.

The pioneering work of Coase (1960) has fundamentally altered the way economists think about externalities. A key insight of his analysis is that all externalities are the result of undefined or poorly defined property rights. The policy prescription seems clear, to alleviate the market failure requires only the assignment of the property right so it can then be priced, and traded, in the marketplace.[30] However, Coase's insight goes farther. As long as the group involved is of small enough number, voluntary bargaining between the parties, without any government involvement, will alleviate the externality.[31] Returning to the earlier example, Burger King could offer to pay McDonald's to stop emitting the air pollution that is interfering in Burger King's production process. Suppose for the sake of example that Burger King would be willing to pay up to $1,000 to stop McDonald's from polluting, while McDonald's could install an antipollution device and eliminate the pollution it emits for $800. As is now well known, the Coase Theorem states that in the absence of significant transactions costs (which would get in the way of the bargaining process), the final allocation of resources will be efficient, and will also be independent of the initial assignment of the right. That is, the same outcome will prevail regardless of whether the government were to intervene and give the right to pollute to McDonald's (in which case Burger King could then offer to pay McDonald's $900 to stop polluting, which they would accept given the antipollution device costs only $800) or if the government were to intervene and give the right to clean air to Burger King (in which case McDonald's would then offer to pay Burger King up to $800 for the right to allow them to pollute, which Burger King would reject, resulting in McDonald's having to install the antipollution device). Since the "high bidder" would be the same in both cases (here Burger King), they would secure the use of the resource and the same outcome would prevail in both cases, and it would be the efficient outcome.[32] However, again, it is important to stress that there is no necessity for the government to intervene to establish the property right because the two firms would have an incentive to bargain out a Pareto-improving solution on their own.

Perhaps the biggest implication of Coase's work is that transactions costs are the fundamental source of unresolved market failures. In cases where a large number of individuals would have to be involved in the bargaining process, high transactions costs might prevent successful bargaining. In the case of large numbers, where bargaining might not occur, the final outcome *will* depend on the initial assignment of property right as it will tend to stay in the hands of the party to whom it was initially assigned.

Based on the insights provided by Coase's analysis, government intervention in the case of externalities when it is warranted (in the case of an unresolved, Pareto-relevant, technological externality) should be limited to establishing or defining private property rights. In some cases, such as the air and oceans, this may not appear feasible, but innovative methods such as tradable pollution permits and tradable fishing rights can accomplish the same task. The modern approach in the case of market failure due to externalities, then, tends to be one of the government creating or establishing more markets (through the defining of property rights), and allowing these markets to work uninhibited, rather than through direct government interventions such as taxes or subsidies along the lines of Pigovian analysis that restricts the role of markets.

7.   MONOPOLY AND ANTITRUST

The next case of market failure to be considered is that of monopoly. It is well established in economics that a monopolist will produce a smaller level of output than the efficient level of output that would be produced under ideal competitive market conditions.[33] This is illustrated in Figure 8 where Q* is the efficient level of market output, Q' is the profit-maximizing output produced by the monopolist, and P' is the profit-maximizing price charged by the monopolist given the firm's marginal revenue (MR) and marginal cost (MC_S) conditions shown.
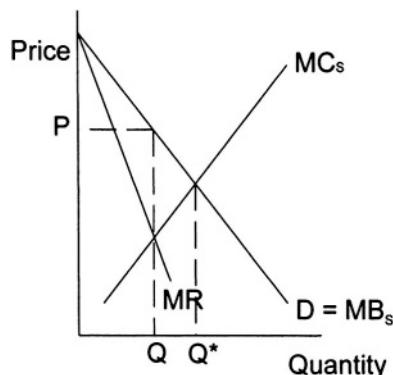

FIGURE 8. Market Failure in the Case of Monopoly.

To understand the proper response of government to a case of monopoly requires making a distinction between two general classes of monopoly—which I shall refer to as artificial monopoly and natural monopoly. An artificial monopoly is a case in which there is only one monopoly firm in the market *because of* an artificial barrier to entry in the industry created by government licensing, law, or regulation. Examples of such cases are the U.S. Postal Service's monopoly on first class mail delivery and local cable providers receiving exclusive contracts from local governments to supply the area with cable service. In these markets, if the artificial barrier to entry was removed, the market would see the entry of new firms to compete with the monopolist. In these cases, it is clear that the market failure itself is due to the preexisting government regulation and that the solution to alleviate the failure is to repeal the law or regulation so that the market is again contestable.[34]

The case of natural monopoly is one in which a single monopoly producer is the natural result of an unregulated competitive market process because a single firm can supply the market at a lower per unit cost than can any combination of smaller firms. Examples of this case may be seen in local markets with a smaller number of consumers, such as one grocery store in a small town or a single newspaper for a small city. However, these examples highlight the extreme difficulty in determining what is, and is not, a monopoly situation because of the somewhat subjective nature of defining the relevant market. While the local newspaper may be the only newspaper, it certainly is competing with many other firms (such as radio and TV stations) in the more broadly defined market of information services. Regardless of whether there are competitors producing closely related goods or services, the most important policy prescription for government in these cases is to ensure that the market remains contestable—that is, that there are no artificial barriers created that would prevent new firms from entering into the industry and competing with the existing monopolist.[35] Just the threat of facing new competition will act as a constraint on the pricing policies of the existing monopolist and lessen the problem of inefficiency.

Traditional economic regulation of a monopoly, either in the form of price regulation or rate of return regulation, is not a very satisfactory solution to the problem of monopoly. Not only do these types of regulation give the monopolist an incentive to misrepresent their true costs and profits, but they generally also result in overcapitalization of assets by the firm.[36] Perhaps more importantly, modern analysis suggests that regulatory agencies tend to get captured by the firms that they regulate and end up working for the interest of the industry at the expense of consumers.[37] Once created, these regulatory agencies may work to help a multi-firm industry act as a cartel, or to help a monopolist maintain its monopoly position in the industry. It is telling along these lines

that once in place, antitrust laws are generally invoked by competing firms, rather than by consumer interests.

If a monopolist firm earns its monopoly status by eliminating competitors through competitive practices like providing consumers with better products at lower cost, it is hard to argue that the monopoly should be subject to government action. Monopolies are rare in the real world, and even the largest giant corporations in the U.S. have a well established history of falling by the wayside. Railroad giants like Norfolk and Western and Union Pacific saw air transportation and trucking evaporate their market; typewriter giants like Royal and Smith Corona were devastated by the introduction of the personal computer; and marketing giants like Montgomery Ward and KMart have fallen to the wayside as modern retailers such as WalMart have provided consumers with better value at lower prices. One only has to look at the high rate of turnover in the Fortune 500 list to know that market power is a temporary phenomenon, at best. The optimal policy for dealing with monopoly power is thus to ensure that markets remain open and contestable so that new firms can, if they wish, enter to compete in the market.

8.   INCOMPLETE INFORMATION

Because information is both costly to provide and to acquire, economic analysis suggest that consumers will generally make decisions with less than perfect information as they economize on their use of scarce resources. It is relatively simple to show mathematically that market situations characterized by less than perfect information are less efficient than outcomes based on full information if one ignores the cost of providing and acquiring the information.[38] Once these costs are factored in, however, an equilibrium with incomplete information may be more efficient than one with full information. Similarly, when government mandates that producers provide certain information to consumers through product labeling or advertising, these policies must be judged by whether the benefits they create outweigh the cost of the additional information. After all, a profit maximizing business firm will sell consumers all the information about the product that they wish as long as consumers are willing to pay a price sufficient to cover the cost to the firm.

It is generally the case that the potential for information problems tends to be more severe for items which consumers purchase infrequently than for items which are purchased on a repeat basis. Consumers not only acquire information through repeated purchases, but the cost to a firm of attempting to take advantage of a consumer is much greater because of the potential for significant losses in terms of lost future repeated dealings with the customer. Information problems thus have the potential to be greater in cases of infrequently purchased items, such as major appliances, or items from souvenir

shops in tourist areas. Economic analysis suggests that brand names are one way in which firms can attempt to provide a quality signal to a consumer for an infrequently purchased item. A traveler stopping at a McDonald's restaurant in Topeka, Kansas, even if he or she has never been to the city before, is ensured a similar quality item to the one provided at the McDonald's restaurant in his or her home town. Furthermore, expenditures on building brand-name capital can be a signal to consumers that the firm is unlikely to be "here today, gone tomorrow" given the large investment expenditure they must recoup.

The most important role for government regarding informational exchanges between buyers and sellers is to provide for a mechanism by which parties can be held liable for making false claims. As long as these mechanisms are in place, the issue is no longer one of the accuracy of information, but of the quantity or quality of the information supplied in the market voluntarily by buyers and sellers. It is also important to note that information can be supplied by outside third-party sources, such as *Consumer Reports* magazine or by producers allowing third-party testing laboratories such as Underwriters Laboratories Incorporated (UL) or Better Housekeeping to test and certify their products.

Markets in which one side of the exchange has more information than the other can be subject to the problem of adverse selection which can destroy the potential for an efficient market outcome. Akerloff's (1970) market for lemons is perhaps the most well-known example of this phenomenon. When the sellers of used cars have more knowledge about the condition of the car than do buyers, the average retail price will reflect the average value of the relative proportions of good and bad quality automobiles in the market. However, at this price a larger number of lower quality cars will be offered for sale (because it is a price above the automobile's true value) while the better quality cars will disappear from the market (because this average price is below the automobile's true value). Similarly in the provision of health insurance, individuals have more information about their potential future health expenses than do the firms providing the insurance. As the average policy price reflects an average, those individuals who expect to incur large future expenditures will choose to purchase insurance, while those individuals who expect to incur small future expenditures will choose not to purchase insurance. This leads to higher average policy premiums that tend to exacerbate this problem. Despite the potential problems in the unregulated market in these cases, the exact role for government intervention is not clear. In the case of health insurance, the government could require everyone to purchase insurance. There are other complications arising from such a policy, however, that might greatly outweigh the benefits. Alternatively, in the case of the used car market the policy prescriptions are even less clear.

The fact that well functioning used car and health insurance markets exist, however, greatly diminishes the relevance of the theoretical results in the cases of incomplete information. As we have previously discussed, it is important to remember that cases of market failure represent cases in which the full gains from trade have not been realized. Thus, cases of market failure represent profit opportunities for entrepreneurs who can, and do, find innovative ways to overcome the sources of the market failure. The development of HMOs, for example, is potentially an example of this type of market innovation in the case of health insurance.

9.   ECONOMIC STABILIZATION AND MONETARY STABILITY

The argument for government intervention to stabilize economic fluctuations over the business cycle is interesting for its lack of philosophical underpinning. Are inflationary booms and economic recessions cases of market failure? Or is it more the case that the stability potentially provided by monetary and fiscal policy can be argued to be a public good that markets cannot efficiently provide? Is a stable monetary environment with low and predictable inflation a public good that can only be efficiently provided by a government or can it be provided through private competing currencies?

Certainly the cases for and against the use of active countercyclical macropolicy are better left for treatment in the field of macroeconomics. The empirical evidence, however, seems to be getting stronger that fiscal policy is not nearly as potent as was once thought in economics under Keynesian models, and that even the best intentioned monetary policy can be destabilizing to an economy due to the timing problems created by lags and the limitations of forecasting.[39] In addition, the insights provided by public choice analysis call into question the ability of the political process to carry out proper fiscal policy. Vote-seeking politicians will generally have an incentive to expand expenditures and cut taxes, and to finance expenditures with debt financing to the greatest extent possible, regardless of the state of the economy. In regard to monetary control, it has long been held that an independent central bank, one removed from the pressures of the political process, will tend to perform better than a politicized central bank.[40] A recent interest has even developed in returning to a system of competing private currencies, rather than government control of the money supply.[41]

Thus, just like in the other cases above, the past several decades have seen a dramatic change away from the view that government intervention in this area is automatically and unquestionably justified, and toward one in which the limitations of real world public sector institutions call for cautious and calculated

intervention only in cases where government can reasonably be expected to actually improve upon the unregulated market outcome.

10.   REDISTRIBUTION

The final area to be explored in this chapter is the role of government in income redistribution. The normal defense of government involvement in this area is on the grounds of equity considerations, making it a more controversial case for intervention than in cases where markets fail to achieve economic efficiency. A notable exception, however, is Hochman and Rodgers (1969) who construct interdependent utility functions across individuals and show that contributions to individuals with lower incomes have the properties of a public good (jointly benefitting everyone through the interdependent utility functions, and not being able to exclude those who don't contribute from enjoying this gain from others' contributions). Using a standard private provision model, they show that the level of contributions to those with lower incomes is less than the efficient quantity. Their results, however, have been met with some controversy as the reality of the assumption of interdependent utility functions is quite arbitrary and lacks empirical justification. In addition, Holcombe and Sobel (2000) argue that interdependent utility functions are precisely equivalent to pecuniary externalities between individuals, and thus they do not create a market failure and require no government correction.

The social contractarian framework also lends itself to a possible justification for redistribution by government.[42] Is it conceivable that at the constitutional decision stage, before everyone knew their future positions in society, that everyone might unanimously agree to put in place a social insurance policy under which those who received the most income would pay taxes that are then transferred to those who receive the least income? If so, then it potentially could be a unanimously agreed upon role of government.

On the other side, arguments against government redistribution can also be made on equity grounds using a procedural theory of fairness, discussed at the beginning of this chapter as standing in contrast to outcome-based theories of fairness. Because a fair outcome is defined as one that is the result of a fair process, it is possible to have outcomes (here income distributions) that are clearly unequal, but are fair nonetheless because they were the result of a fair process (the market allocation mechanism). One could apparently counter this with an argument that the market allocation mechanism is a unfair process. Again, because we are dealing with an issue of subjective value judgements, there is very little room for objective science to help settle this dispute. Nonetheless, following the original line of reasoning, the forceful redistribution of wealth by government may be thought of as an unfair process

because of its coercive nature. If so, then any outcome of this forceful redistribution must necessarily be unfair regardless of the equality present in the final outcome.

Even if one agrees that there is a role for government in redistribution, there is still a lack of agreement about the degree or extent of the redistribution because of the lack of a positive, objective definition of equity. Furthermore, the greater the extent of the redistribution, the larger will be the distortions and movements away from efficiency in the markets that are taxed to provide the funding for the redistributive activities. Perhaps most compelling is the fact that any attempt to redistribute wealth or income through the public sector will necessarily alter the incentive to produce, not only for those taxed in order to finance the transfer, but also for those receiving the transfer benefits. It is impossible to use market prices to efficiently allocate resources, communicate information, and motivate economic participants without also relying on those prices to determine the distribution of income.[43]

Finally, it is worth discussing whether real world political institutions are (1) more efficient than private firms at providing redistribution, and (2) capable of directing the payments toward those individuals who need it most, rather than to those with the most political influence. Because public sector redistribution crowds out private sector redistribution, it is unclear exactly how much private sector charity there would be in the absence of government involvement. Going back in history to the early 1900s, prior to the U.S. federal government's involvement in redistribution to the extent it is today, most adults were members of private mutual-aid societies. Members joining one of these "clubs" contribute and when anyone in the club was in need, other members would provide assistance. The extent of fraudulent claims was vastly lower in this private system than it is today in the public sector welfare system because the members generally all knew one another.

Recent events after the September 11, 2001 terrorist attacks on the World Trade Center provide some additional evidence on private charitable giving. Rather than proposing a massive government redistribution scheme, President George W. Bush on national television called for individuals to voluntarily contribute to private charities that provided assistance to those who were affected. In response, within five weeks after the attack, 70 percent of Americans had reported giving some type of charitable support (58% reported giving money, 13% blood, and 11% time donations). By the end of November, less than three months after the attacks, relief organizations had raised over $1.1 billion in voluntary donations.[44] The massive outpouring of private volunteers who gave their time and labor, as well as those who made financial contributions is substantial evidence that in cases where redistribution is widely deemed as appropriate, that it will be given in generous quantity. The massive extent of charitable giving after the World Trade Center attack would seem

to be evidence against the formal model presented by Hochman and Rodgers (1969) in which charitable giving is virtually impossible to provide through private markets due to the public good nature of the contributions.

Modern public economic research focuses less on exploring the merits of, or the optimal conditions for, redistribution and rather is more focused on attempting to explain the patterns of actual redistribution that occur. In the United States, for example, only about one-sixth of all transfers are means tested (that is, the qualifications for receiving the transfer are dependent on income). The fact that many redistribution programs tend to benefit middle income households, or large organized industries, is not surprising from the standpoint of public choice theory. First, because winning the vote of the decisive median voter is of critical importance for securing electoral victory, one might predict that transfers would be taken from both the upper and lower tails of the income distribution and targeted at the middle.[45] Secondly, concentrated interest groups will always have an advantage at securing transfers from widespread and unorganized groups who do not have the political power to oppose the redistribution. Subsidies to operas and home mortgage interest deductions seem to be two examples of redistribution clearly not aimed at the lower end of the income distribution.

While the justification for government intervention in the case of redistribution is subject to much debate, the fact is that modern democratic governments generally devote more than half their budgets toward transfer activities. Evidence suggests, however, that these transfers are captured by those groups with political influence, rather than those most in need. Because government redistribution crowds out private charities that are more effective at directing the payments to those most in need, it is potentially the case that transfers to those most in need could be increased by reducing or constraining the role of government in redistributive activities.

11.   CONCLUSION

This chapter has summarized the cases for and against government intervention into markets to improve social welfare, either through increasing economic efficiency or equity. Beginning in the late 1800s through the mid 1900s, there was rapid development of very rigorous neoclassical economic theory to these cases, and founded upon this analysis was a presumption that government intervention could automatically be used to solve most of these problems. The downfall of Keynesian macroeconomic theory coupled with the development of public choice theory in the late 1900s, however, has shifted the tide somewhat. Modern analysis incorporates the idea that real world political institutions, just like markets, are subject to failure. In many cases of market failure, the best policy will be that of no policy because government intervention is likely to result in an even more inefficient outcome than is already

present. The late twentieth century has seen a dramatic evolution from an era in which the mathematical proof of market failure was a sufficient condition for government intervention to one in which it is not.

ACKNOWLEDGMENT

NOTES

1. For an insightful analysis of the constraints imposed on the U.S. government by the U.S. Constitution relative to the constraints imposed by the Articles of Confederation that proceeded it, see Holcombe (1991).
2. Bator (1958) is widely cited as a fundamental paper listing the cases in which market failure is likely.
3. Quote taken from Stigler (1993), p. 402.
4. For similar academic arguments along these lines see Buchanan (1962) and Buchanan (1975).
5. Buchanan (1962) discusses the implications of this difference for the potential of government action to improve on inefficient market outcomes.
6. See Mueller (1989), Chapter 18, and Munger (2001) for discussions of the problems with voting.
7. Readers interested in a more formal treatment of the conditions necessary for economic efficiency in a competitive general equilibrium (Arrow-Debreu) framework are referred to Myles (1995), Chapter 2. A nice concise graphical interpretation can be found in Cullis and Jones (1998), Chapter 1. The notion of Pareto optimality was first derived in Pareto (1909). Little (1959), however, was the first person to name the condition as such.
8. The original development of this criterion can be found in the works of Hicks (1940) and Kaldor (1939). Cullis and Jones (1998), Chapter 2, contains a nice review of this criterion as well as the later improvements to the criterion introduced by Scitovsky (1941).
9. It should be noted that economics generally makes the assumption that each individual is the best judge of his or her own welfare (or utility) and that social welfare may be captured as simply a sum (or weighted sum) of the welfare of the individuals that make up the society.
10. Baumol (1982) for example, attempted to define a fair outcome as one in which there was an absence of envy. While intuitively appealing, Holcombe (1983, 1997) illustrates several cases in which an outcome is envy free, but it is clearly not fair. Thus even the absence of envy does not imply fairness.
11. This procedural theory of fairness is generally associated with Rawls (1971) and Nozick (1974) and is widely applied in the field of constitutional economics. For general overviews of this field see Gordon (1976) and Buchanan (1990).
12. For an overview of several functional forms see Cullis and Jones (1998), Chapter 1. A more in depth analytical treatment may be found in Varian (1978), Chapter 1, and Heathfield and Wibe (1987), Chapter 5.
13. It is again worth pointing out that this condition is necessary for the efficient provision of any good that is joint-in-consumption, regardless of whether it is excludable or nonexcludable. Only in the case where the good is additionally nonexcludable is it a public good

subject to the potential for market failure due to the free-rider problem. Goods that are joint-in-consumption but are excludable are often called club goods, and while there is the potential for some consumers to be inefficiently excluded from consuming the good under private provision, Buchanan (1965) shows how clubs can arrive at the optimal production of the good. In his model, often called "the theory of clubs," the optimal sharing group (club size) and optimal quantity of the good produced are simultaneously determined. The optimal club size will be finite as long as the good is subject to congestion.

14. Readers interested in the mathematical derivation of the private provision equilibrium are referred to Myles (1995)), Chapter 9, or Cornes and Sandler (1996), Chapter 6. While the standard Nash equilibrium outcome in this private provision model produces an outcome in which the public good is undersupplied, other characteristics of this equilibrium do not seem to fit real world data and experimental evidence very well. Because of this, models with alternative conjectural formulations other than Nash have been developed, but have still not proved very satisfactory.

15. Klein (1987) provides a nice examination of how tie-in sales can allow markets to efficiently provide public goods.

16. See Cowen (1988) for a comprehensive examination of the many critiques of standard market failure arguments such as this.

17. See Holcombe and Sobel (1995) for evidence on this point. Their paper also contains an example of a widely used empirical model that is useful for estimating the degree of publicness a good exhibits.

18. The median voter outcome is sometimes called Bowen equilibrium and is generally attributed to the work of Bowen (1943). Hotelling (1929), Downs (1957), and Black (1958) also made important contributions to median voter theory.

19. If decisions were subject to a unanimous voting rule, however, Lindahl prices for every individual would create unanimous agreement at the efficient output level. Wicksell (1896) was a famous proponent of the use of the unanimous decision rule for collective choice, and the statement above is sometimes more formally stated as Lindahl prices create Wicksellian unanimity at the efficient output level. For a more in depth discussion of the relationship between the median voter model, Lindahl prices, and Wicksellian unanimity see Holcombe (1985).

20. In addition, as Denzau and Mackay (1976) show, Lindahl pricing can result in outcomes that seem rather odd from an equity standpoint. For example, to finance the provision of a radio transmission tower (where the height of the tower or strength of the signal was the good in question), the person with the highest marginal benefit from expanding the quantity of the good (and thus the person with the highest tax share) would be precisely the person in the group with the weakest signal that would be improved by the additional production. The person with the strongest signal, living next door to the tower, would have a Lindahl tax price of zero as they gain no marginal benefit from additional provision of the good.

21. See Mueller (1989), Chapter 14, for a summary of the overwhelming empirical evidence on this point.

22. The applicability of the bureaucracy model as a model of real world outcomes remains a controversial issue. Niskanen himself has acknowledged the limitations of this model, see Niskanen (2001). For evidence against the applicability of the simple bureaucracy model see Bohm (1987) and Jackson (1982).

23. See Breton and Winetrobe (1975) for the analytical treatment of the equilibrium size of a budget-maximizing bureau.

24. In fact, it would require that the median voter's tax share was significantly larger than his or her benefit share (or equivalently his or her tax price was larger than the Lindahl tax price)

such that with the expansion in the quantity produced under bureaucratic supply, that QB would equal Q* even though the median voter's most preferred quantity, QM, was less than both.

25. The terminology that distinguishes pecuniary from technological externalities was first used by Scitovsky(1954).

26. Buchanan and Stubblebine (1962) were the first to formally note the distinction between inframarginal and Pareto-relevant externalities.

27. See Pigou (1924). Baumol (1992) contains an excellent review of optimal Pigovian tax policy in the case of a negative externality.

28. The government, however, should not use the revenue collected from this tax to compensate those suffering from the external cost because it would not give other individuals the appropriate disincentive to avoid suffering the cost. For example, if subsidies were paid to compensate owners of houses near airports for the noise they suffer, there would be less incentive to avoid building houses near the airport. If compensation were paid, more houses would locate near the airport, increasing the external cost per takeoff.

29. Note, however, that to generate the revenue required to grant the subsidy would require imposing a tax in another market which, except in the case of a lump-sum tax, would create an additional distortion in the economy.

30. While property rights to resources should be clearly defined for market efficiency, Holcombe and Sobel (2001) show that individuals should not be allowed to claim ownership rights to the *value* of the resources they own. Establishing rights to the value of resources internalizes pecuniary externalities and results in markets moving away from efficiency rather than toward it.

31. A famous example of this is the case of the spillover that exists between apple growers and honey-producing beekeepers, that was originally cited by Meade (1952) as a case of a technological externality that would result in market failure as not enough beekeepers would locate next door to apple growers as would be efficient. Cheung (1973), however, found that in the state of Washington, there was a long history of contractual arrangements in which beekeepers were paid for their contributions to apple growing.

32. The traditional illustration of the Coase Theorem as presented here ignores any income effects that result from the establishment of the property right. Even if income effects are considered, an efficient outcome will prevail, but it will be a different efficient outcome. To illustrate, imagine that the two cases correspond to two different points in an Edgeworth box, both of which are off of the contract curve. In both cases, bargaining will lead to a Pareto optimum along the contract curve, but which efficient outcome emerges will depend on the starting point.

33. Here I give the treatment of monopoly less attention than the cases of public goods and externalities. This relative weighting is traditional in the field public economics as monopoly, and the regulation of monopoly, are often covered in more detail in the field of industrial organization. A reader interested in a more in depth treatment of these issues is referred to Tirole(1988).

34. A contestable market is one in which it is relatively costless for new firms to enter into the market to compete with existing sellers.

35. Splitting a natural monopoly into several smaller firms would be an unwise policy choice because it would lead to several smaller firms, each with a higher cost of production than the single large firm.

36. This overcapitalization by a firm under rate-of-return regulation was first shown by Averch and Johnson (1962), and is known as the Averch-Johnson effect.

37. Readers interested the capture theory of regulation and papers dealing with the problems of traditional economic regulation are referred to Stigler (1971), Posner (1975), Peltzman (1976), and Benson, Greenhut, and Holcombe (1987).

38. Like monopoly, incomplete information is a subject generally relegated to the field of industrial organization, so here I only treat it in a cursory manner. Again, a reader interested in a more in depth analysis is referred to Tirole (1988).

39. See Rasche and Thornton (2001) and Gwartney, Stroup and Sobel (2000), Chapter 15 for evidence along these lines.

40. See Toma (2001), Alesina and Summers (1997), and Eijffinger and Schaling (1995) for a discussion of and evidence on central bank independence and economic performance.

41. See Solomon (1996), Craig (1996), and Good (1998) for additional information about competing currencies and private money.

42. This body of literature explores the evolution of constitutions (which are known in this literature as social contracts) and is also known as the field of constitutional economics. For a general overview see Gordon (1976) and Buchanan (1990). The idea of redistribution as a preconstitutional social insurance scheme was first developed in Buchanan and Tullock (1962).

43. A strong argument along these lines is made in Chapter 1 of Friedman and Friedman (1980).

44. Data is from "A Survey of Charitable Giving After September 11th, 2001" undertaken by the Independent Sector, October 23, 2001 available at http://www.independentsector.org/PDFs/Sept11_giving.pdf and from Robert A. Sirico, "Charity Bill Would Expand Private Gifts," *The Grand Rapids Press,* November 28, 2001 available at http://www.acton.org/research/editorials/sirico/charitybill.html.

45. This theory is sometimes called Director's law of income redistribution (named after Aaron Director who proposed it), and an exposition of it can be found in Stigler (1970). Tullock (1971) also presents a similar argument.

REFERENCES

Akerloff, George A. (1970). "The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84 (3): 488-500.

Alesina, Alberto and Summers, Lawrence H. (1997). "Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence." In *Independent Central Banks and Economic Performance,* Eijffinger, Sylvester C. W., ed. Cheltenham, U.K. and Lyme, N.H.: Edward Elgar.

Averch, Harvey and Johnson, Leland L. (1962). "Behaviour of the Firm under Regulatory Constraints." *American Economic Review* 52 (5): 1052-1069.

Bator, Francis M. (1958). "The Anatomy of Market Failure." *Quarterly Journal of Economics* 72 (3): 351-379.

Baumol, William J. (1992). "On Taxation and Control of Externalities." *American Economic Review* 62 (3): 307-322.

Baumol, William J. (1982). "Applied Fairness Theory and Rationing Policy." *American Economic Review* 72 (4): 639-651.

Benson, Bruce. (1990). *The Enterprise of Law: Justice Without the State.* San Francisco: Pacific Research Institute for Public Policy.

Benson, Bruce, Greenhut, M. L., and Holcombe, Randall. (1987). "Interest Groups and the Antitrust Paradox." *Cato Journal* 6 (3): 801-817.

Black, Duncan. (1958). *The Theory of Committees and Elections.* Cambridge: Cambridge University Press.

Bohm, Peter. (1987). *Social Efficiency,* Second Edition. London: Macmillan.

Bowen, Howard R. (1943). "The Interpretation of Voting in the Allocation of Economic Resources." *Quarterly Journal of Economics* 58 (1): 27-48.

Breton, Albert and Winetrobe, Ronald. (1975). "The Equilibrium Size of a Budget Maximizing Bureau." *Journal of Political Economy* 83 (1): 195-207.

Buchanan, James M. (1962). "Politics, Policy, and Pigovian Margins." *Economica* 29: 17-28.

Buchanan, James M. (1965). "An Economic Theory of Clubs." *Economica* 32 (125): 1-14.

Buchanan, James M. (1975). "Public Finance and Public Choice." *National Tax Journal* 28 (December): 383-394.

Buchanan, James M. (1990). "The Domain of Constitutional Economics." *Constitutional Political Economy* 1 (1): 1-18.

Buchanan, James M. and Stubblebine, Craig, W. (1962). "Externality." *Economica* 29 (116): 371-384.

Buchanan, James M. and Tullock, Gordon. (1962). *The Calculus of Consent.* Ann Arbor: University of Michigan Press.

Cheung, Stephen N. S. (1973). "The Fable of the Bees: An Economic Investigation." *Journal of Law and Economics* 16(1): 11-34.

Coase, Ronald H. (1960). "The Problem of Social Cost." *Journal of Law and Economics* 3: 1-44.

Coase, Ronald. (1974). "The Lighthouse in Economics." *Journal of Law and Economics* 17 (2): 357-376.

Cornes, Richard and Sandler, Todd. (1996). *The Theory of Externalities, Public Goods, and Club Goods,* Second Edition. Cambridge: Cambridge University Press.

Cowen, Tyler (ed.) (1988). *The Theory of Market Failure: A Critical Evaluation.* Fairfax, Virginia: George Mason University Press.

Craig, Ben. (1996). "Competing Currencies: Back to the Future?" Federal Reserve Bank of Cleveland, *Economic Commentary.*

Cullis, John and Jones, Philip. (1998). *Public Finance and Public Choice.* Second Edition. New York: Oxford University Press.

Denzau, Arthur T. and Mackay, Robert J. (1976). "Benefit Shares and Majority Voting." *American Economic Review* 66 (1): 69-76.

Downs, Anthony. (1957). *An Economic Theory of Democracy.* New York: Harper & Row.

Eijffinger, Sylvester C. W. and Schaling, E. (1995). "The Ultimate Determinants of Central Bank Independence." In *Positive Political Economy: Theory and Evidence,* Eijffinger, S. and Huizinga, H., eds. New York: Wiley & Sons.

Friedman, Milton and Friedman, Rose. (1980). *Free to Choose: A Personal Statement.* New York: Harcourt Brace & Company.

Good, Barbara A. (1998). "Private Money: Everything Old is New Again." Federal Reserve Bank of Cleveland, *Economic Commentary.*

Gordon, Scott. (1976). "The New Contractarians." *Journal of Political Economy* 84 (3): 573-590.

Gwartney, James D., Stroup, Richard L., and Sobel, Russell S. (2000). *Economics: Private and Public Choice,* Ninth Edition. Fort Worth, Texas: The Dryden Press, Harcourt College Publishers.

Hayek, F. A. (1945). "The Use of Knowledge in Society." *American Economic Review* 35 (4): 519-530.

Heathfield, David F., and Wibe, Sören. (1987). *An Introduction to Cost and Production Functions.* London: Macmillan.

Hicks, John R. (1940). *Value and Capital.* Oxford: Clarendon Press.

Hochman, Harold M. and Rodgers, James D. (1969). "Pareto Optimal Redistribution." *American Economic Review* 57 (4): 542-557.

Holcombe, Randall G. (1983). "Applied Fairness Theory: Comment." *American Economic Review* 73 (5): 1153-1156.

Holcombe, Randall G. (1985). *An Economic Analysis of Democracy.* Carbondale: Southern Illinois University Press.

Holcombe, Randall G. (1991). "Constitutions as Constraints: A Case Study of Three American Constitutions." *Constitutional Political Economy* 2 (3): 303-328.

Holcombe, Randall G. (1996). *Public Finance: Government Revenues and Expenditures in the United States Economy.* Minneapolis/St. Paul, MN: West Publishing.

Holcombe, Randall G. (1997). "Absence of Envy Does Not Imply Fairness." *Southern Economic Journal* 63 (3): 797-802.

Holcombe, Randall G. and Russell S. Sobel. (1995). "Empirical Evidence on the Publicness of State Legislative Activities." *Public Choice* 83 (1/2): 47-58.

Holcombe, Randall G. and Sobel, Russell S. (2000). "Consumption Externalities and Economic Welfare." *Eastern Economic Journal* 26 (2): 157-170.

Holcombe, Randall G. and Sobel, Russell S. (2001). "Public Policy Toward Pecuniary Externalities." *Public Finance Review* 29 (4): 304-325.

Hotelling, Harold. (1929). "Stability in Competition." *Economic Journal* 39: 41-57.

Jackson, Peter M. (1982). *The Political Economy of Bureaucracy.* Oxford: Philip Allan.

Johansen, Leif. (1963). "Some Notes on the Lindahl Theory of Determination of Public Expenditures." *International Economic Review* 4 (3): 346-358.

Kaldor, Nicholas. (1939). "Welfare Propositions and Interpersonal Comparisons of Utility." *Economic Journal* 49 (195): 549-552.

Klein, Daniel. (1987). "Tie-ins and the Market Provision of Public Goods." *Harvard Journal of Law and Public Policy* 10: 451-474.

Lindahl, Eric. (1919). "Just Taxation — A Positive Solution." reprinted as in *Classics in the Theory of Public Finance,* Musgrave, R. A. and Peacock, A. T., eds., pp. 168-176. London: Macmillan, 1958.

Little, I. M. D. (1959). *A Critique of Welfare Economics.* Oxford: Clarendon Press.

Meade, James E. (1952). "External Economies and Diseconomies in a Competitive Situation." *Economic Journal* 62 (245): 54-67.

Mueller, Dennis C. (1989). *Public Choice II: A Revised Edition of Public Choice.* Cambridge: Cambridge University Press.

Munger, Michael C. (2001). "Voting." In *The Elgar Companion to Public Choice,* Shughart II, William F. and Razzolini, Laura, eds., Chapter 9, pp. 197-239. Cheltenham, UK: Edward Elgar.

Myles, Gareth D. (1995). *Public Economics.* Cambridge: Cambridge University Press.

Niskanen, William A. (1968). "Nonmarket Decision Making: The Peculiar Economics of Bureaucracy." *American Economic Review* 58 (2): 293-305.

Niskanen, William A. (1971). *Bureaucracy and Representative Government.* Chicago: Aldine-Atherton.

Niskanen, William A. (2001). "Bureaucracy." In *The Elgar Companion to Public Choice,* Shughart II, William F. and Razzolini, Laura, eds., Chapter 11, pp. 258-270. Cheltenham, UK: Edward Elgar.

Nozick, Robert. (1974). *Anarchy, State, and Utopia.* New York: Basic Books, 1974.

Pareto, Vilfredo, (1909). *Manuel d'économie Politique.* Paris: Girard and Briere,

Peacock, Alan. (1979). *The Economic Analysis of Government and Related Themes.* Oxford: Martin Robertson.

Peltzman, Sam. (1976). "Toward a More General Theory of Regulation." *Journal of Law and Economics* 19 (2): 211-240.

Pigou, Arthur C. (1924). *The Economics of Welfare,* Second Edition. London: Macmillan.

Posner, Richard A. (1975). "The Social Costs of Monopoly and Regulation." *Journal of Political Economy* 83 (4): 807-827.

Rasche, Robert H., and Thornton, Daniel L. (2001). "The Monetary/Fiscal Policy Debate: A Controlled Experiment." *Monetary Trends,* p. 1. Federal Reserve Bank of St. Louis.

Rawls, John. (1971). *A Theory of Justice.* Cambridge, Mass.: Belknap.

Samuelson, Paul A. (1954). "The Pure Theory of Public Expenditure." *Review of Economics and Statistics* 36 (4): 387-389.

Samuelson, Paul A. (1955). "Diagrammatic Exposition of a Pure Theory of Public Expenditure." *Review of Economics and Statistics* 37 (4): 350-356.

Scitovsky, Tibor. (1941). "A Note on Welfare Proposition in Economics." *Review of Economic Studies* 9(1): 77-88.

Scitovsky, Tibor. (1954). "Two Concepts of External Economies." *Journal of Political Economy* 62 (2): 143-151.

Smith, Adam. (1976) [1776]. *An Inquiry into the Nature and Causes of the Wealth of Nations,* Edwin Cannan, ed. Chicago: University of Chicago Press.

Solomon, Lewis D. (1996). *Rethinking Our Centralized Monetary System: The Case for a System of Local Currencies.* Westport, Conn.: Praeger.

Stigler, George J. (1970). "Director's Law of Public Income Redistribution." *Journal of Law and Economics* 13 (1): 1-10.

Stigler, George J. (1971). "The Theory of Economic Regulation." *Bell Journal of Economics and Management Science* 2(1): 3-18.

Stigler, George J. (1993). "Monopoly." In *The Fortune Encyclopedia of Economics,* David R. Henderson, ed. New York: Warner Books.

Tirole, Jean. (1988). *The Theory of Industrial Organization,* Cambridge, Mass.: MIT Press.

Toma, Mark (2001). "Monetary Policy." In *The Elgar Companion to Public Choice,* Shughart II, William F. and Razzolini, Laura, eds., Chapter 21, pp. 440-456. Cheltenham, UK: Edward Elgar.

Tullock, Gordon. (1971). "The Charity of the Uncharitable." *Western Economic Journal* 9: 379-392.

Varian, Hal R. (1978). *Microeconomic Analysis.* New York: W.W. Norton.

Wicksell, Knut. (1958) [1896]. "A New Principle of Just Taxation." reprinted as pp. 73-118 in *Classics in the Theory of Public Finance,* Musgrave, R. A. and Peacock, A. T., eds, London: Macmillan, 1958.

Chapter 3

FISCAL CONTITUTIONALISM

Geoffrey Brennan
The Australian National University
Geoffrey.Brennan@anu.edu.au

Giuseppe Eusepi
University of Rome "La Sapienza"
eusepi@dep.eco.uniroma1.it

Abstract	The object of this paper is to provide a general overview of that area of intellectual activity characterised by the application of so-called "constitutional economics" to fiscal phenomena. We call this area "fiscal constitutionalism." In modern times, it has been associated most notably with the work of James Buchanan and those operating in his tradition (among whom we more or less count ourselves).
Keywords:	Constitutional economics, James Buchanan, fiscal constitutionalism
JEL classification:	H20, H50

The object of this paper is to provide a general overview of that area of intellectual activity characterised by the application of so-called "constitutional economics" to fiscal phenomena. We shall call this area "fiscal constitutionalism." In modern times, it has been associated most notably with the work of James Buchanan and those operating in his tradition (among whom we more or less count ourselves).

Like most areas of scholarship, fiscal constitutionalism has been developed partly by the application of intellectual logic and partly by accident. It reflects, necessarily, the particular interests and intellectual orientation of Buchanan himself—and the different interests and orientations of the scholars that have subsequently taken up and developed the Buchanan approach. Buchanan, though originally a public economist, has a broad range of intellectual interests in political philosophy and much of his work is addressed to fundamental questions in political philosophy. This strand of his work invites

comparison with the work of Hobbes and Locke, of Hume and Kant, of Rawls and Gauthier. It would in principle be possible to examine fiscal constitutionalism through the lens of political philosophy in this way—and to identify the specifically *fiscal* elements as an application of Buchanan's larger project in constitutional theory. Seen in such terms, two questions would be invited: First, how is Buchanan's constitutional theory to be located in relation to other variants of contractarian political theory [Hobbes and Locke say] and to criticism of the contractarian approach, such as Hume's? Second, how does the specifically fiscal element in Buchanan's account connect to his broader constitutionalism?

We mention these questions here at the outset to indicate an approach to our subject matter that we shall *not* be taking. Here, we shall approach fiscal constitutionalism from the "other end" as it were. We shall focus on the ways in which fiscal constitutionalism contrasts with standard "public economics". This strategy has the virtue that it more closely corresponds with the actual history. Fiscal constitutionalism developed as a reaction to and critique of orthodox public economics; and in to a significant extent the fiscal applications pre-dates the development of broader political/philosophical themes in Buchanan's work. In that sense, it makes as much sense *historically* to see Buchanan's broader constitutional project as an outgrowth of the fiscal one, rather than the fiscal elements as an application of more general constitutional theorising. In any event, we shall here finesse the large political questions and direct attention instead to the less abstract world of taxes and expenditures and debt and to how these phenomena are most appropriately understood and analysed.

1.   THE CONSTITUTIONAL APPROACH

Our point of departure is then to draw a contrast between the orthodox approach to fiscal questions [henceforth 'OA'] and the constitutional approach [henceforth 'CA']. In drawing that contrast, it is useful to begin by describing OA briefly. Consider what is probably the most familiar normative question within the OA tradition: how should the tax system be designed so that it best achieves certain goals, deemed appropriate to tax systems? The goals in question would normally be objectives like "horizontal equity" and "vertical equity," and "efficiency." Analysis of how alternative tax systems score with respect to the various goals might be pursued separately or simultaneously. So a familiar formulation of one aspect of the problem could be something like: "how should the tax system be designed to secure maximum efficiency?" This problem would normally be understood as requiring us to specify how the tax system should be organised so as to secure a given and independently determined amount of revenue at least aggregate cost to taxpayers. Included

specifically in that cost would be the welfare losses that taxpayers sustain in substituting less valued for more valued activities/items-of-consumption as a result of "distortions" introduced by the tax system. Another related question might be: "how should the tax rate structure be designed so as to secure the best compromise between efficiency (so understood) and vertical equity—where 'vertical equity' is conceived as the requirement to allocate the tax burden appropriately across individuals with different levels of appropriately measured 'well-offness'? So questions would arise here as to how 'well-offness' should be appropriately conceived;"[1] and once appropriately conceived, how best measured; and how the distribution of effective burdens under alternative tax regimes translates into the distribution of well-offness, so measured. And so on.

Clearly, the OA in its public-economics/public-finance guise is no different from what is done in the normative analysis of economic policy issues more generally. A problem presents itself—unemployment; significant balance of payments deficits; pollution; rising crime rates;—and it seems utterly natural, and indeed totally unexceptionable, to ask how the problem is to be solved; and to conceive of the solution in terms of specifying the most appropriate values of available policy instruments. Put a little more generally, the appropriate object of direct normative concern is the particular *policy* that government might enact.

The constitutional approach—CA—is to be distinguished from OA in this particular respect. In the CA view, the appropriate domain of normative theorising is not policies themselves but rather the elements of the institutional framework under which the policies are determined. CA takes as its point of departure a particular challenge to OA—what traditionally has been cast as the "benevolent despot" challenge. That is, according to the CA critique, OA implicitly assumes a benevolent despot conception of politics. The "despot" element arises because policies are assumed to be available for direct choice by policy-makers without the mediation of any political constraints. The "benevolence" element arises from the fact that the policy-makers are assumed to use their discretion to act according to the normative considerations that the OA economic analysis proffers.

Both the benevolence and the despotism aspects of OA are taken by CA critics to be objectionable. CA insists that the political element, as a matter of fact, always enters into the determination of policy outcomes, even in those cases where a more or less single "policy-maker" can be identified. [To see that this latter aspect is non-trivial, consider the case in which a policy decision—whether to introduce a new tax or not—is to be determined by plebiscite. In such a case, it is misleading to think of policy as being chosen by any identifiable individual to whom the economist might proffer advice. The policy choice simply emerges from the process that aggregates the various views of

all the enfranchised persons. There is, so CA insists, simply no escape from the political element. Certainly no escape in broadly democratic regimes, but probably not in more dictatorial ones either, though in the latter case the political constraints would, of course, take a rather different form.

In the discussion that ensues, we shall assume a democratic political setting. Within such a setting, any analysis of policy determination that omits the role of electoral constraints is to that extent defective. Now, as all economists know, specification of all the relevant constraints is a necessary element in all policy analysis, whatever its ultimate purpose. If the objective is strictly positive/explanatory—if, that is, we seek to explain why a particular policy regime is as it is—we had better take account of the fact that policy "choosers" are political agents who want to win office. And that in order to secure office, those policy choosers have reason to give voters what they want.[2] To seek to explain what policy instruments are actually chosen on the basis of what is "best" according to certain independently derived normative criteria, seems on its face to be an inadequate model of the policy determination process. At the very least, one would seem obliged to explain why it is that the "best" policies are likely to be the ones chosen. Equally, if the object of the analysis is normative—that is, if the object is to secure "improvements" in the state of the world by policy means—it can hardly be appropriate to ignore operative political constraints. Moreover, there is a rather deeper normative objection. Suppose it were an assumption of analysis that policy choosers were totally benevolent, then either the democratic political constraints would be irrelevant (encouraging policy-makers to choose what they would choose anyway) or those constraints would prevent policy-makers from "doing good." That is, democratic institutional constraints would be presumptively undesirable. On this reading, then, OA is "anti-democratic."

The anti-democratic edge to OA is not a mere fancy. Consider the following example. A Royal Commission is appointed to examine the tax system and to make recommendations for reform. The Commission appointed happens to be composed of public finance academics of the orthodox kind who deliberate on the basis of the standard normative theory of tax design and come to a consensus as to the best (feasible) tax system; and duly publish their findings. The resultant recommendations are then filtered through the houses of parliament where the package of tax measures is picked over, various bits removed and other bits added. The public finance community, previously exultant at the sweep and coherence of the recommended package, is now outraged at the intervention of all these expedient know-nothings, wheeler-dealers and political mountebanks. "If only," they moan, "the politicians had kept out of it for once." Clearly, there is implicit here a distinct elitist element—one that sits uncomfortably with democratic sympathies. We do not necessarily claim that the economist experts are wrong to complain here. But we do think there is an

issue that the OA tends to suppress. There are, in summary, serious objections to the "despot" assumption on both positive and normative grounds. What of the "benevolence" aspect?

The CA critique here is an important and well-known element in the development of "public choice" political theory. Simple extrapolation from behaviour in market contexts seems to require that political actors—whether in their roles as voters, or politicians, or bureaucrats—will exhibit essentially the same motivational structure as do market actors. Simply to assume that actors in their political roles will operate totally benevolently while in their market roles those same actors will operate as egoistic wealth maximisers seems both implausible as a matter of fact and objectionable as a matter of ideological neutrality. That is, the "two-hats hypothesis" embedded in OA seems to pre-suppose a kind of schizophrenia that at the very least requires some justification. And this not least because the hypothesis assigns to political institutions specifically some strange capacity to purify the motives of human actors. Making this point does not require us to make extreme assumptions about the total egoism of actors in their market roles; we can allow that some element of benevolence, or more general desire to act as morality requires, is part of agents' motivational structure—in both political *and* market arenas. What the public choice model insists upon is the imposition of *motivational* symmetry across all institutional structures; and the implication of corresponding *behavioural* symmetry, except where reasons grounded in the theory of rational behaviour can be provided to indicate otherwise.

As against the implicit "benevolent despot" element, characteristic of OA, CA seeks to embed policy determination within an explicit model of political process; and specifically within a model that is obedient to standard economists' assumptions used in the analysis of markets. What this means is that policy instruments themselves are not available for direct "choice"—whether choice based on the dictates of normative criteria or on the particular preferences of the policy chooser. Rather, policies must be interpreted as emerging from a process—as aspects of a political equilibrium, much like prices in a general equilibrium model of markets. On this view, "choosing" a policy is very like consumers "choosing" a price in the market for widgets. Recall that the critique of "just price" theory is that it applies normative criteria to the "wrong" variable. Just price theory attends to a particular price when it is the whole market process that requires evaluation. In the same way, CA accuses OA of attending to the wrong variable. On the CA view, the proper domain of normative evaluation is the set of political institutions from which policy outcomes emerge. Once it has been judged that a particular set of such institutions is the best feasible, then the policy regimes that emerge from those institutional arrangements have to be identified as the best *feasible*—where feasibility now includes explicitly the process within which the policy regime is determined.

2.   FISCAL CONSTITUTIONS

The kinds of questions that would arise under the CA rubric as described above might include the following. What policies are likely to arise under the kinds of political institutions that are most common in Western democracies? How are such policies likely to differ from those emergent under other less democratic regimes? What are the likely policy consequences of altering political institutions in particular ways?

Such questions are unlikely to be able to be answered in any great detail. The analysis will not enable us, in general, to predict who will win the next election. Or whether this or that product is likely to be exempted under the indirect tax system. Or whether the defence base will be located in one city rather than another. Or just how large the level of spending on education or health will be. The insights if any that the analysis will be able to provide will be at a much broader level of abstraction. So, for example, on such issues as: the tendency for certain regimes to rely more heavily on debt or inflationary financing than on current taxation; or the propensity for democratic regimes to use regulatory rather than directly budgetary measures in particular policy areas; or the overall distributional impact of fiscal operations; or the regional distribution of fiscal benefits. Attempts might be made to "explain" certain aspects of particular policies by reference to "public choice considerations." For example, one might seek to explain why firms in industries with lots of "high-risk" jobs might lobby more extensively for publicly provided worker-insurance schemes. But the limits of such exercises must be acknowledged. Though any such explanation represents an interesting and useful illustration of public choice methods, it does not explain why one particular policy is implemented and another with similar redistributive properties fails. As Jonathan Pincus puts it in an early piece of empirical work in public choice analysis [Pincus (1977)], one may be able to explain why tariff protection as a general phenomenon has significant electoral attractions without being able to explain satisfactorily why product A receives tariff protection while B does not. In short, the public choice analysis of policy is much better at explaining the general structure of policy outcomes than the specific content of prevailing policy regimes.

The agenda of questions laid out above describes what we might call the fiscal aspects of a general political science. It focuses on the budgetary policy implications of particular political arrangements, with an eye to assessing the details of those arrangements. Compared with ordinary political science therefore, it is distinctive in respect of the particular subset of political attributes it focuses on—viz on pieces of institutional detail—and in respect of the particular subset of consequences that it attends to—viz the policy outcomes.[3] Particular research questions might include, for example, the effects of term

limits on debt financing; or of federal structures on the size of the public sector; or of the Presidential veto on the extent of geographical special interest spending. What is at stake in each case is the analysis of the mapping from institutional arrangement to general policy consequence. The ultimate normative ambition is to make recommendations as to the "best" form of institutions; but one might reckon the enterprise of charting the connections interesting in its own right and potentially useful, without signing on to any such large-scale normative ambition.

3.   TAXES AS PRICES

None of what we have said so far, however, captures a particular theme that has been significant in the public choice/public finance tradition—namely, the idea that tax instruments might be better understood as themselves *institutions* rather than as policy consequences. On this reading, there is something special about the tax system—something that makes it logically prior to other policy decisions, and indeed embeds the tax system within the broad structure of the constitution, broadly construed. This distinctive feature of the tax system is one thing that scholars have had in mind in referring to the fiscal constitution and it is therefore important to examine it here. That is the aim of this section.

3.1.   The Wicksell-Lindahl Tradition

Although the idea of the tax system itself as an institution within the political system can be generalised across a variety of political models, the point of departure for that idea historically lies in the interpretation of taxes as a kind of political equivalent to market prices. That interpretation is an explicit piece of the analysis in the influential work of Knut Wicksell[3] and in the exposition of Wicksell's ideas by his student Erik Lindahl [Lindahl (1919)].[4] Wicksell's interest in this issue had been piqued by earlier work on the distribution of the tax burden across income classes. Wicksell appears to have been concerned that the aristocracy was enjoying the major share of public expenditure benefits, while the working class was paying the bulk of taxes. At the same time, he also seems to have been concerned with the possibility that the rising tide of democracy might serve to transform the tax system into one that effectively taxed only the rich. His anxieties in this latter connection seem to have been focussed not only on the distributive consequences of such possibilities per se. Wicksell seems also to have been concerned about the effects of such tax arrangements on the possibility of exploitative public expenditure programs-as the dominant working class voted for expenditures for which it did not have to pay.

Wicksell's concern was to devise political arrangements so that such "unjust" outcomes were precluded—hence his insistence on the rule of unanimity

for any new expenditure activities. Within this scheme, taxes would be endogenous. Any expenditure would require tax financing that secured the agreement of [virtually] all parties; and this in turn would imply that those who saw themselves as benefiting from a particular public program would have to pay for that program in additional taxes. This idea of taxes as analogous to market prices enters explicitly in Lindahl's attempted formalisation of Wicksell's political model. The political bargaining process that Lindahl envisages has the property that in equilibrium taxes are such that burdens are borne in relation to marginal benefit from the public good supplied. Although Lindahl conjectures that this outcome would in practice be replicated by a proportional income tax, there is no implication that tax arrangements should be set independently of the political process itself. Indeed, any prior restrictions on tax arrangements, of the kind characteristic of conventional public finance, can only serve to prevent mutually beneficial tax-expenditure combinations from meeting the unanimity test. In other words, there is no "fiscal constitution" as such in the Lindahl/Wicksell model—just a requirement of unanimity and a deliberate absence of further restrictions. With these requirements in place, the relevant classes [and interests of other kinds] are free to bargain. This bargaining process under [virtual] unanimity is the Wicksell/Lindahl picture of political process.

3.2.   Majority Rule

Actual democratic political processes are however not characterised by unanimity, even the "virtual unanimity" that Wicksell conceded would be necessary. And once one moves away from the unanimity requirement to the practically more familiar case in which political decisions are taken by some form of majority rule, then political bargaining clearly cannot be relied on to limit political exploitation. It is conceivable that majorities will form to vote into effect special interest expenditures *and* special interest tax packages [i.e., ones in which the burdens are focused on the minority]. Indeed, it seems extremely *likely* that this kind of majority exploitation of minorities will occur, because under any form of political competition, policy packages will tend to emerge that focus maximal benefits on those whose votes are sought.

One means of limiting the scope for such exploitation under majority rule might be to impose the requirement of broadness of tax base, more or less along the lines of traditional "horizontal equity requirements." Clearly, a taxation system that requires the costs of any public expenditures to be distributed evenly across individuals in relation to income or aggregate consumption, say, restricts—though it does not fully eliminate—the scope for fiscal exploitation. Majorities may still vote for special benefit expenditures, even if they are to be financed by general taxes.[5] Nevertheless, a horizontal equity constraint on the

tax side may be one aspect of a "second-best" institutional arrangement if the "first best" (virtual) unanimity arrangement is ruled out for other reasons. This will be a "second-best" outcome not just because it insulates the citizenry from endless rounds of exploitative taxation, but also because it sets limits to the levels of taxation that will be politically demanded. There is, in other words, an explicit recognition here of the fact that when the majority has to pay taxes for the special interest legislation it demands, the levels of such special interest expenditure will be lower—lower, that is, than if the cost-share/tax-price that majority members face had been zero.

The resultant defense of horizontal equity in the tax structure depends as much on its effects on the "efficiency" of the collective decision-making process as on the properties of the tax system as such. In that respect, the restrictions have a "constitutional" aspect: they are part of the "rules of the political game." Such tax restrictions operate a little like constitutional restrictions on the *taking* powers of governments or on the capacity of governments to introduce retroactive legislation. All such restrictions recognise that ordinary majority rule can be consistent with outcomes that we seek explicitly to rule out. But tax restrictions are more "process-oriented" than mere outcome restrictions, because they identify tax arrangements as having an influence on the nature of collective decision making across the board. In the case of fiscal constitutions, we are seeking not just to limit the range of political outcomes directly [by ruling some out]; we are also seeking to make the political process itself work "better." In that sense, tax restrictions are more like bi-cameralism, or federal structures, or various types of separation of powers than they are like entrenchment of particular policy outcomes: these tax restrictions constitute part of the political process itself.

4.   PUBLIC FINANCE IN CONSTITUTIONAL PERSPECTIVE—ANALYSIS

Within the perspective of the fiscal constitution, what particular aspects of the tax system invite special analytical scrutiny? How in particular does the kind of analysis of tax instruments in this CA perspective differ from that in the more familiar OA? Clearly, the most significant difference is that whereas OA focuses attention on the effect of tax changes on the composition of private goods consumption, with aggregate revenue fixed, CA focuses analysis on the public goods/private goods margin—that is, on the effect of tax changes on the *level* of expenditures. Within the standard public choice model of political process, the issue of exactly how alternative tax instruments translate into 'supply conditions' for public goods is the crucial aspect.

In what follows, we aim to illustrate the kinds of considerations at stake in this translation exercise. The object is to provide a feel for how analysis might

go, without any attempt to be exhaustive. We gather the illustrative points we wish to make under various heads, reflecting different relevant aspects of the "cost-share" approach. Throughout the ensuing discussion, we shall assume that revenues are to be used to finance a pure public good of the Samuelsonian type. This means specifically that the good is available to all in equal amounts—with the valuations that different individuals place on the good in question at various levels known and fixed. This simplification allows us to focus on the "supply" side with the demand side appropriately simplified.

4.1.   Cost-share Determinacy

It is clear that, if taxes are to operate as quasi-prices for public goods, voter-taxpayers must know in advance what the cost shares are. Each must be able to calculate the cost to herself in terms of extra tax dollars of an additional quantum of public expenditure. This requirement involves not merely the prior specification of the tax instrument in question, but also the specification of taxes in a manner that allows translation into cost-share form. Suppose we are as taxpayers considering some particular expenditure program. Suppose all that we know about the tax arrangements is the prevailing tax system, with its particular mix of taxes. Suppose each knows also for each tax what s/he will pay for a given change in the tax rate and what revenue that rate increase will secure. Each can on this basis calculate the proportion of the revenue from that tax source that s/he supplies [on average]. But clearly, s/he would need to know what specific tax or mix of taxes is to be used before s/he could determine what the relevant cost share actually was. On this basis, cost-share determinacy seems to require earmarking of taxes for specific purposes; or alternatively, an assurance that as revenue levels expand, the proportions in which each tax contributes to revenue remain the same or changes in some fashion that is specified ex ante.

In fact, we have simplified the calculations necessary by assuming that changes in aggregate revenue are secured solely by *rate* changes. Revenue can be altered as well by altering the tax base, sometimes in rather subtle ways, and voters would need to know ex ante, even for an earmarked tax, whether additional revenue would be secured by a rate increase or a base change. Moreover, in talking of a "rate increase," we are presupposing that there is only one rate applied. But, clearly, under progressive rate structures there are many rates, any one or more of which might in principle be raised to secure additional funds. Unless the degree of progression is held constant so that all rates rise pari passu, or the particular rate changes that are to be increased are specified ex ante, the cost share for any particular voter will remain indeterminate.

In short, the determinacy requirement provides presumptive arguments for earmarking (or tax-mix constancy), and for either single rate ["proportional"]

rate structures or rate structures of the "degressive" kind [a uniform flat rate with a specified and fixed exemption]. One way of achieving uniform rates is to use commodity taxes which are typically of that form; but flat rate direct taxes are no less acceptable. There is a strong connection between the determinacy requirement and the age old demand for tax system "simplicity;" but some apparently simple tax systems do not permit citizens readily to determine their cost shares. Conceptual clarity requires that determinacy be given independent status in the "fiscal constitution." In fact, there is a conceptual connection between the cost-share determinacy notion and the requirement that no tax be retrospective. A retrospective tax is also one to which taxpayers cannot respond appropriately. However, a tax can satisfy the non-retrospectivity requirement but fail the cost-share determinacy test. Non-retrospectivity effectively requires that the tax liability be determined by reference only to market choices undertaken *after* the tax arrangement is specified; cost-share determinacy requires a similar temporal pattern in respect of voters' *political* choices.

4.2. Cost-share Estimation—Margins versus Averages

It may seem as if the determination of cost-shares is simply a reformulation of traditional OA concerns with tax incidence and distribution. Estimation of one's cost share will after all require each to calculate how much tax she pays in comparison with the average person [or equivalently in proportion to total revenue]. So all the complications of tax incidence and the like are involved no less in the CA approach. Adding a model of the political process to distributional calculations may be required; but the familiar OA concerns all reappear in a modified guise. So much is clearly right as far as it goes. But it should not lead us to conclude that there is nothing more required. One additional element in the CA analysis of tax, for example, is the distinction between "average" and "marginal" incidence—between the distributional effects of a tax in toto and the distributional attributes of the revenue *increments* from that tax.

A simple example may illuminate here. Consider a progressive income tax. A characteristic feature of such a tax is that, considering revenue-take as a whole, burdens are distributed across taxpayers so that those taxpayers with larger tax base [let it be income, somehow measured] pay proportionately more tax. Often the maximum rates of tax imposed on the richest are limited by "Laffer curve" considerations; and it is self evident that there is at least an upper bound on such tax rates of [somewhat less than] 100%. Suppose the rate structure in a particular instance is [0; 15%; 30%; 50%] with the transition points at specified income levels. To preserve that rate of progression—to maintain the same cost-shares across income classes—if revenue demands were to double would require that the rates double, with no change in the transition points. The required rate structure would be [0; 30%; 60%; 100%];

which is infeasible at the top end. Hence, it is almost axiomatic that, in such a case, the additional revenue will be distributed in a fashion less progressive than the total burden. And indeed, in many cases, under a progressive rate structure, the actual cost shares for marginal expenditures can be extremely regressive. Again to take a simple example, a flat tax with an exemption of say $8000 and a flat rate of 30% will raise additional real revenue if there is inflation and the real value of the exemption is reduced thereby. That extra revenue is, however, effectively a lump sum tax on all persons with incomes over $8000; the marginal cost share is not even proportional to income—it is the same for virtually all.[6] So a provision that retains the progressive structure in toto involves a highly regressive structure "at the margin"—with margin here understood not in terms of formal "marginal tax rates" but in terms of increments to revenue. Whereas the OA focuses on the trade-off between more income and more leisure, CA focuses on the trade-off between more or less public expenditure levels. There are different 'margins' at stake; and it may require some effort of imagination for someone versed in the OA to translate standard tax analysis into the public goods-private goods dimension of choice relevant to political decision-making.

4.3.   Excess Burdens in Constitutional Perspective

In the same spirit, excess burdens of taxes make an appearance in the CA approach, but now in terms of the aggregate cost of public expenditures—rather than as mere efficiency losses to be treated as normatively relevant in their own right. To provide for an expenditure of $100m will cost taxpayers more than $100 by virtue of the excess burden generated through the taxing process [as taxpayers substitute, say, less valued leisure for more valued goods in response to the tax]. That much is totally standard. But what is of interest in the collective decision-making context is the way in which these excess burdens increase as revenue increases. Thus, an emergent literature on 'the marginal cost of public funds' [see, for example, Browning (1974)] suggests that the cost of an incremental public expenditure of say $1 m, will be very much higher than $1 m, and much more than casual inspection of the size of the "Harberger triangles" would suggest. Perhaps more relevant, the presence of such marginal excess burdens, in any political process in which voters' preferences have some influence, will tend to reduce the level of public spending [*ceteris paribus*]. This means in turn that the measurement of the efficiency gains of particular tax substitutions using the standard "equi-revenue" approach characteristic of public finance orthodoxy will always involve underestimation, because it will ignore the effects on the level of public spending itself.

One striking example of the role of excess burdens in a political model is the simple model of the 'welfare state' first elaborated by Buchanan (1975/2001)

and Meltzer and Richard (1981). In that model, revenue is assumed to be de-
rived from a proportional income tax and is used to finance a simple lump-sum
transfer to all voter-taxpayers. Each voter is presumed to vote according to the
net benefit received from the fiscal transaction. Accordingly, in the absence
of excess burdens, any voter with less than average income will want all of
GDP distributed through the fisc; and any voter with more than average in-
come will want none of GDP so distributed. Since the income distribution
is skewed towards the lower end, there will be more of the former than the
latter—or equivalently the decisive "median voter" will have an income less
than the average—and so the whole of GDP will be channelled through the fisc
in political equilibrium. The inclusion of excess burdens from the tax-transfer
process moderates this extreme knife-edge result. The size of the revenue take
will be determined by the condition that the median income-earner's "cost-
share" *plus the median income-earner's marginal excess burden* be equal to
marginal transfer benefit. The efficiency effects of the tax-transfer mechanism
will be reflected both in total income (and hence the amount to be redistributed
through the fisc) and in the median-income-earner's utility as s/he substitutes
out of higher valued goods into lower valued leisure in response to the income
tax. Both effects tend to moderate the size of the revenue take—and corre-
spondingly, the transfer each receives. In this sort of model, the inclusion of
excess burdens is crucial in achieving plausible results; and in particular it is
the effect of those excess burdens on the rational calculus of taxpayer-voters
that is relevant, rather than the normative significance of excess burdens in
themselves.

   Of course, if there were independent reasons for thinking that the public
sector was over-expanded, then it is conceivable that the movement to a less
efficient tax regime might increase efficiency overall. That is one result that
emerges from the Leviathan model of government analysed in some detail
by one of us in another place and another collaboration [see Brennan and
Buchanan (1980a, b)]. In that model, the preferences of citizen-voters count
only at the constitutional level in the choice of alternative tax regimes. Aggre-
gate public spending is given by the maximum revenue that government can
extract from the constitutionally assigned tax instruments. In other words, the
Leviathan model involves no connection between voter preferences over pub-
lic goods and spending levels. In the Leviathan model, accordingly, cost-shares
as such are irrelevant. "Mixed models" in which both demand-side influences
from voter preferences and supply-side influences from a quasi-monopolistic
government remain to be developed. In both demand-driven and supply-driven
models, however, the efficiency properties of the tax system are significant. In
the demand-driven case, the influence of excess burdens operates via net voter
demands for public spending. In the supply-driven case, excess burdens gen-
erated under different tax regimes are a proxy for the total revenue that can be

obtained under those regimes and hence the level of public spending that will emerge.

4.4.   Fiscal Illusion

Within the OA tradition, fiscal illusion receives relatively little attention. Understandably so, because it can be reasonably assumed that rational agents have an incentive to inform themselves about those aspects of the tax regime that bear on their market choices. So, for example, if there is an excise tax on beer, I as a taxpayer need to know what the gross (cum tax) price of beer is in order to make rational choices between beer consumption and other activities. But I do not need to know how much of the gross price of beer is made up of tax. If there were an increase in the price of beer, my response in the consumption of beer would be (asymptotically) the same whether that price increase was attributable to a tax increase or had occurred by virtue of a hop crop failure. In choosing among different jobs, I will need to know the return net of income-tax; and this may require me to calculate my average tax rate—and also my marginal tax rate if I can adjust hours of work in one or both employments. If there is any complex calculation to be made in this connection, then my market choices may be more liable to error; and this risk would constitute an argument for simpler rather than more complex tax arrangements—an argument based on "fiscal illusion" considerations. But in the OA tradition, fiscal illusion is relevant only in so far as it bears on market choices.

Within the CA tradition, by contrast, fiscal illusion is a much more central issue—partly because the knowledge requirements are much more substantial in relation to political choices and partly because the incentive to acquire the relevant information is radically reduced. Clearly, if I am to make intelligent political decisions about the general range and extent of government activity, I need to have a good sense of just how much I am paying in taxes. I need to know what proportion of the gross price of beer—and gasoline and tobacco products (and other goods often subject to excise)—is attributable to tax and what to costs of production. I need to know how much income tax I pay; how much corporate tax; and so on. And I need to know this not just in terms of the formal wedge between cum-tax and net-of-tax magnitudes but in terms of the actual incidence of the tax and expenditure in question. That is, I need to make a comparison between the situation with the tax-and-expenditure operation and without it. Needless to say, such comparisons are often a complicated matter requiring some considerable expertise to unravel. They are in many instances subject to contention even among the so-called experts.

Further, no voter has much incentive to acquire that information. As the literature on "rational ignorance" shows, the fact that no voter can reasonably expect to be decisive means that the free-rider problem that is the basis of political provision in the first place is replicated in relation to information about

public goods. If I make a mistake in voting and vote for the wrong option, that mistake will almost certainly exercise no influence on the prevailing political outcome. This fact means that truly rational voters will be '(rationally) under-informed' [Downs (1956)] and may have negligible incentive to vote their interests even if they could perceive those interests accurately [Brennan and Lomasky (1993)]. Why should a voter undertake the tedious calculations required to exercise a fully informed vote? Why should s/he even become informed about what the relevant rates of tax actually are? Only, presumably, if s/he derives pleasure from knowing the right answers for their own sake.

If there is little incentive for the voter to do any really hard work in discerning the truth about tax arrangements, there is no less incentive for rival political parties to disguise the costs (and advertise the benefits) of the policies that they propose. Under almost any political regime, it will pay candidates to choose tax arrangements—other things equal—that involve substantial fiscal illusion. There is a natural centrifugal force inclining tax systems toward extensive reliance on those tax instruments of which taxpayers are least aware. Even parties who might stand for reduced public spending will have an incentive to background taxes. Consider, for example, just such a party contemplating a tax and expenditure cut. The party has a choice between two taxes to abolish. One is a tax of which taxpayers are highly conscious; the other a tax of which they are hardly aware. To make the biggest impact and reap the largest electoral rewards from the taxpaying constituency, the best tax to cut is the most conspicuous. Obversely, a party committed to public sector expansion will tend to choose relatively invisible taxes to expand—and focus public attention as best it can on the virtues of the proposed expenditures.

It seems a necessary property of political equilibrium that the tax system that emerges will be one crafted, as if by an 'invisible hand' one might say, to be maximally invisible. Opposition parties/candidates will certainly have an incentive to disclose the true cost of opponents' policies. But even here there is a trade-off between such disclosure and advertising the merits of their own alternative expenditures, financed to a significant extent out of revenue sources that are just as invisible as those the rival is using. To blow the whistle on the true cost of one's rival's expenditures is to expose the true cost of one's own. Of course, there are some offsetting considerations—the perceived fairness of taxes; and the capacity of alternative instruments to raise the required revenue. It is not, for example, an electoral asset to be *seen* to be trying to deceive one's constituents—so that occasionally even a display of apparently electorally risky candour can be advantageous. Nevertheless, there are strong systematic forces encouraging reliance on less visible tax instruments: fiscal illusion is likely to be an endemic problem in political systems of all stripes. Democratic systems specifically are not exempt from this problem.

When we talk of "fiscal illusion" here, we have in mind a broader range of policy instruments than merely budgetary ones. In fact, it seems likely that policies implemented through the budget will be more conspicuous than those implemented through regulatory and other "off-budget" devices. An explicit subsidy to taxi cab drivers/owners in the budget is more conspicuous than a restriction of entry to the taxi industry, though the benefits to taxi driver/owners may be equivalent. We say "may be equivalent" here because once fiscal illusion effects are allowed for, ceteris are not (politically) paribus. The greater conspicuousness of explicit budgetary devices implies that the effective subsidy is likely to be much larger where the policy is implemented through regulatory processes.[7]

If there is a systematic bias in political process towards invisible taxes and visible expenditures, as the foregoing observations suggest, then there is a corresponding tendency towards over-expansion bias for the public sector as a whole. In recognition of that bias, citizens at the constitutional level may well want to insist that taxes satisfy some reasonable overtness requirement and that there be limits on the extent to which policies can be pursued outside the budgetary process. Although such restrictions may be interpreted as anti-democratic, in the sense that they stand in the way of what ordinary majority rule politics is likely to produce, it seems clear that in another sense they are provisions designed to help majority rule work better. Such restrictions serve to limit the natural tendencies of political processes to keep voters in the dark.

To illustrate how all this bites in relation to OA, consider the question of 'optimal commodity taxation' under the two approaches. It is now well accepted that to achieve efficiency in commodity taxation in the choices among private goods, the ideal (feasible) tax regime will not involve uniform taxes but rather taxes with differential rates according to the degree of substitutability with necessarily tax-exempt leisure. This is one of the now standard 'optimal tax theory' results.[8] So while OA endorses a general presumption of broad-based commodity taxation, there is certainly no uniform rate requirement and indeed quite elaborate rate variations will typically be required if the tax system is to be maximally efficient in OA terms (ie minimise excess burdens). But the CA case for strictly uniform rates stands. That case depends much more on considerations of simplicity, and minimisation of fiscal illusion—considerations that relate to the electoral choice between public and private goods—than on raising revenue at the least conceivable cost. The CA analyst will point to the difficulties for taxpayers in assessing what they are actually paying, and the scope for political exploitation and for fiscal smokes and mirrors, once the principle of tax rate heterogeneity is accepted. If the efficiency losses attributable to distorted private goods choices under a uniform rate regime were large enough, then CA would have reason to pause. But the CA conviction is that efficiency in the overall public-private choice is a much more significant issue

and that considerations that work towards that *political* dimension of choice, such as simplicity and intelligibility of the tax regime, should be dominant. Here, then, as elsewhere, OA and CA diverge in respect of policy implications as well as in respect of the focus of argument.

4.5.   Debt Financing

One specific application of the CA, that has been of considerable interest to CA scholars for much of the last half-century, is the question of debt financing. Approaching debt from a Wicksellian perspective, one is immediately confronted with the scope for "exploitation" of future generations by virtue of the fact that those future generations are not directly enfranchised in the political bargaining process. If future generations cannot veto expenditure proposals that they will have to pay for, then a putative unanimity rule offers no protection. The only way in which protection might be afforded is to constitutionally ban funding mechanisms that impose the primary burden on future generations. Whether debt does in fact impose the burden of current spending on future generations has been a contested matter in the economics profession over the last fifty years. Buchanan's first major book (1958/1999) was devoted to demonstrating that debt financing did indeed involve passing the burden of expenditure to future taxpayers. His target in that book was the proposition that, because the resources used up in any current expenditure are necessarily current resources, there can be no *real* inter-temporal shift in funding. There can be no net addition to the nation's current resources, unless the borrowing is external. Accordingly, on this view, there is a radical divide between internal and external public debt: the former we "owe to ourselves:" the latter we "owe to others."

It is not necessary here to retrace Buchanan's arguments against this general view. But it is worth noting that what is at stake is Buchanan's charge that it involves an inappropriate degree of aggregation. The "we" referred to needs to be disaggregated into various sets of persons—current taxpayers, current bond-holders, future taxpayers—before meaningful analysis can begin. Further, one needs to distinguish between fisc and nation. For it is clear that the taxing power of any government falls considerably short of command over all of the nation's income. The power to borrow adds to the rise's capacity to extract revenue, even from *current* citizens. But current citizens only give up those additional resources because future taxpayers will be obligated to repay the debt with interest.

More recently, another challenge to the CA anxiety has emerged, associated with a "rediscovery" of the so-called Ricardian equivalence theorem. The claim here is effectively that future generations do not need to be present in order to be protected, because provided all members of the present generation make positive bequests, present voters will rationally treat one dollar of

additional future taxes as equivalent to one dollar of present consumption forgone at the margin.[9] In this connection, much of the relevant argument has hovered around the issue of "debt illusion." However, as many commentators, following Ricardo's original formulation in fact, have argued, although the Barro claim may be correct under conditions of full information, in practice taxpayer-voters are simply not fully aware of the future tax implications of current debt-financing decisions. Such lack of awareness may not matter hugely in itself, but given the incentive that politicians have to minimise the apparent cost of public spending measures, debt illusion is likely to be subject to systematic exploitation. And here as the CA perspective emphasises, the critical issue is the effect that debt financing has on the level of public spending. There is a clear empirical hypothesis at stake here: Debt financing leads to higher public spending than tax financing *ceteris paribus.* And an important element in the explanation of why this is so relates to "debt illusion."[10]

5.   FISCAL CONSTITUTIONS—PROCESS

It is a theme in CA analysis that normatively driven conclusions must be institutionally feasible. That is, there must be some way for changes deemed normatively desirable to be implemented under plausible conditions. We might ask, for example, why the analysis in the foregoing section could not be incorporated into a conventional approach. Why could not the effects of particular tax choices on the conduct of political processes be added to the standard treatment as a significant addition without any particular assault on the OA framework? Put another way, how is it that the recommendations of CA analysis in respect of effects on public sector size whether via cost-share or fiscal illusion mechanisms, do not themselves fall subject to the benevolent despot critique? In particular, if recommendations about alternative taxes on conventional efficiency grounds are suspect, why are not recommendations about alternative taxes based on their capacity to produce efficient political outcomes similarly suspect?

Part of the answer to this set of questions is that there is nothing necessarily wrong with the normative recommendations of OA, providing they can be reinterpreted and given institutional expression within a constitutional framework. Specifically, if we can conceive the taxation system as being chosen as part of an appropriate 'constitutional exercise', and nested within a choice context where constitutional aspects seem likely to be paramount, then we can allow the recommendations of OA to play whatever role they can within that setting.

What might such a constitutional setting look like? Suppose, to take a not too fanciful example, that the choice of the taxation system is to be lifted out of

day-to-day political process and assigned to a body of appropriately represen-
tative persons who are themselves reasonably expert in taxation matters. Sup-
pose that the tax system so chosen is to come into effect some years hence—
say, five years; and suppose further that it will once settled on be in place for
a further extended period—say, fifteen years. Suppose the persons chosen for
the representative committee are academics and lawyers who are constrained
by professional norms to choose according to criteria that at least purport to
reflect the "public interest." Suppose finally that the recommended tax system
will have to be accepted or rejected by the government of the day as an all-or-
nothing package. Failure to accept the recommendations will commit the fisc
to the prevailing tax system for another five years at which point a further such
Taxation Enquiry can be initiated.

   We observe that this scenario is "not too fanciful" because it resembles a
practice that is familiar in the Westminster democracies [U.K., Canada, Aus-
tralia, etc.]. The "Royal Commission" or "Committee of Enquiry" of the kind
typified by the Carter Commission in Canada or the Asprey Committee in
Australia is essentially a quasi-constitutional body of the type described in the
previous paragraph. Such Committees operate in a context where tax change of
a broad systemic kind is rare and where the changes (if any) once made are ex-
pected to be in place for a generation and perhaps longer. Being divorced from
immediate political/electoral concerns, the members of such Committees are
free to devise a tax system that they believe to be defensible on the basis of
broad issues of principle. Of course, they are also free to indulge themselves
with every eccentric whim and fiscal fancy that their imaginations can sup-
ply. But assuming that they do not wish to appear objects of public ridicule,
they are likely to be constrained to produce a set of recommendations that will
command tolerable broad respect. Often enough, as in the Carter Commission
case, there is a good supply of orthodox academic public finance; and issues
like horizontal and vertical equity and efficiency will typically be in evidence
as explicit criteria to be applied in the exercise. Indeed, if there is a systematic
fault in the outcomes of such "Enquiries" it is as much that "academic" points
in tax design are likely to be excessively entertained and feasibility constraints
not adequately recognised, as that there will be evidence of special interest
bias. Sometimes, parts of such Reports read rather like academic treatises, as
if the whole were written more to impress one's academic colleagues than to
design a tolerably well-working revenue system. Here though, the institutional
setting seems conducive to a general constitutional approach. And the content
of the recommendations seems to be uncannily similar to the familiar recom-
mendations of OA. Arguably, the recommendations place too little emphasis
on the considerations of determinacy and simplicity and overtness and polit-
ical viability that we discussed in the previous section. But, against the spirit
of Wicksell and Lindahl, there does seem to be something politically desirable

about having tax institutions selected quite deliberately outside the context of
ordinary electoral politics. And institutionally, that is what the Committee of
Enquiry device seems to do.

Interpreted in this way, there does not seem to be anything inherently ob-
jectionable about applying normative criteria directly to tax instruments. The
chief problems from a constitutional perspective seem rather to revolve around
the possibility that tax "reform" might be excessively frequent; or that tax
changes occurring outside the quasi-constitutional setting may be too partial
or too much influenced by the special interests of currently dominant majori-
ties. It is, in short, part of the logic of the constitutional approach that the
processes through which tax changes are made are important elements in the
content of the tax policy recommendations. On the CA view, attention should
focus as much on those processes as on the recommendations themselves.

6.   SUMMARY AND CONCLUSIONS

If you were to inquire of the passer-by in the street whether your nation
should have a constitution, you are likely to get a response something like:
"Of course! What a stupid question!" It is just taken for granted that every
self-respecting country has to have some kind of constitution, even if (as in
Britain's eccentric case) this constitution is not actually written down any-
where. But if you asked that same passer-by whether the country ought to
have a specifically "fiscal constitution," you would be more likely to get a re-
sponse of blank amazement. When you explained that what you have in mind
is that governments should be prevented from running deficits or using debt
financing in other than extreme circumstances, and that taxes should not be
allowed to be levied retrospectively, and that there might be limits placed on
total tax take; or that changes in the basic tax system should be decided through
processes that are quasi-judicial rather than party-political in character; or that
there might be some general requirement that the tax system be intelligible in
its effects and clearly apparent to taxpayers—then perhaps the passer-by might
think you were onto something important. You might be accused of being anti-
democratic; or of not showing enough respect for the President. But when you
point out that the same response would be no less apt in relation to more famil-
iar constitutional provisions which the passer-by has just, in principle at least,
endorsed, and that we need clear and effective tax arrangements if democracy
is to work well, the passer-by might grudgingly concede that maybe you have
a point.

However, the passer-by is not the only person you have to convince. The
relevant audience is as much the public finance expert—whose intellectual dis-
positions incline him to think of public economics in a rather different way. As
Buchanan shrewdly observes in the preface to his book *Public Finance in De-
mocratic Process* (1967), there is at stake in the constitutional approach a shift

in perspective very like that involved in simple visual illusions. Such a shift in perspective requires an exercise of the imagination; and one that comes hard to those whose views are already fixed along particular lines. A first step in this shift involves the incorporation of political 'feedback loops' in the analysis of particular tax changes. That step seems relatively easy for the OA mind, partly because it involves no major assault on the basic normative framework. But to conceive fiscal phenomena as part of the basic institutional structure of society—much like federalism or the separation of powers—and to evaluate them accordingly involves an important additional step, and is the characteristic feature of the 'constitutional approach'.

What we have tried to suggest in this essay is something of what is at stake analytically in that approach. The argument has been cast almost exclusively within a demand-driven model of political processes where the general cast of results is most similar to those of orthodox public finance. Arguments for generally broad-based uniform taxes emerge much like in orthodox discussion, though the rationale for such taxes is rather different and the claims of strict uniformity of rates much stronger than in the "optimal tax" informed orthodoxy. If the political process is seen to be dominated by "supply-driven" considerations, so that political outcomes reflect much more the preferences of politicians and bureaucrats than of ordinary citizen-voters, then the policy thrust under the constitutional approach can diverge quite markedly from that under conventional public finance—an aspect explored in some detail in Brennan and Buchanan (1980a, b). This fact simply serves to underline the central relevance of political assumptions. In the OA framework, political considerations are totally ignored. In the CA framework, they are placed center stage.

NOTES

1. This issue bears on questions that go beyond the design of tax systems and that have a life in literature outside public economics. For more 'philosophical' discussion of some aspects of the same issue, see for example, Cohen (1990); Dworkin (1981a, b); Fleurbaey (1995); and Sen (1992).

2. Specifying precisely what it *is* that voters want is a critical issue in any theory of democratic processes. It is also a matter of some contention within CA circles, for reasons that we shall take up below.

3. The influence that Wicksell's (1896) habilitation thesis had on James Buchanan, following a fortuitous discovery in the stacks of the University of Chicago library is legendary. It is also worth noting that Samuelson's influential analysis of public goods derived from a reading of Musgrave's (1938) treatment of the Lindahl model. By a variety of separate routes the Wicksell-Lindahl tradition has deeply penetrated contemporary public finance, though the constitutionalist element has not been a significant part of that legacy.

4. The earlier work of Mazzola in applying the marginalist approach to public goods should be acknowledged here. See Mazzola (1890/1958).

5. In recent work, Buchanan and Congleton (1998) have applied the idea of horizontal equity more generally across every element of the budget under the rubric of a 'generality principle'. This application illustrates the attributes of a fiscal constitution perhaps more strikingly. And here, very much in the Wicksellian spirit, the object is to achieve via restrictions on majoritarian processes, the outcomes that would emerge from an idealised unanimity collective decision rule.

6. Those with incomes below $8000 post-inflation will, of course, still have a cost share of zero.

7. This is the point that Buchanan and Tullock (1975) exploit in explaining why environmental policies are more commonly pursued via regulatory than budgetary means. Regulated industries prefer the regulation to taxation because regulation can increase profit. And the environmental lobby prefers regulation because it recognises that pollution restrictions will be larger if the industry does not have to pay. Regulation is the result of a mutually beneficial bargain between environmentalists and regulated firms at the expense of consumers of the regulated product and general taxpayers.

8. See, for example, Atkinson and Stiglitz (1980) Lecture 2.

9. This way of presenting the argument is a gloss on the currently fashionable version, for which see Barro (1974). It does no serious violence to Barro's position in our view.

10. Though not the only element. In Brennan and Buchanan (1980a, b), we show that illusion is not a necessary condition for debt to reduce future incomes even when all agents make positive bequests.

REFERENCES

Atkinson, A. B. and Stiglitz, J. (1980). *Lectures in Public Economics.* New York: McGraw-Hill.

Barro, R. J. (1974). "Are Government Bonds Net Wealth?" *Journal of Political Economy* 82: 1095-1118.

Brennan, G. (1972). "Second Best Aspects of Horizontal Equity Questions." *Public Finance/Finances Publiques* 3: 282-291.

Brennan, G. and Buchanan, J. M. (1980a). *The Power to Tax.* Cambridge: Cambridge University Press.

Brennan, G. and Buchanan, J. M. (1980b), "The Logic of the Ricardian Equivalence Theorem." *Finanzarchiv* 38: 4-16.

Brennan, G. and Lomasky, L. (1993). *Democracy and Decision. The Pure Theory of Electoral Preference.* Cambridge: Cambridge University Press.

Browning, E. (1974). "The Marginal Cost of Public Funds." *Journal of Political Economy* 84: 283-298.

Buchanan, J. M. (1958). *Public Principles of Public Debt.* Homewood: Richard D. Irwin.

Buchanan, J. M. (1960). " 'La scienza delle finanze:' The Italian Tradition in Fiscal Theory." In *Fiscal Theory and Political Economy.* Chapel Hill: University of North Carolina Press.

Buchanan, J. M. (1967). *Public Finance in Democratic Process.* Chapel Hill: North Carolina University Press.

Buchanan, J. M. (1975/2001). "The Political Economy of the Welfare State." Collected Works of James M. Buchanan, Vol. 13, *Politics as Public Choice.* Indianapolis: Liberty Fund Inc.

Buchanan, J. M. (1975). *The Limits of Liberty: Between Anarchy and Leviathan.* Chicago: University of Chicago Press.

Buchanan, J. M. (1976). "Barro on the Ricardian Equivalence Theorem." *Journal of Political Economy* 84: 337-342.

Buchanan, J. M. (1993), "How Can Constitutions Be Designed So That Politicians Who Seek To Serve 'Public Interest' Can Survive and Prosper." *Constitutional Political Economy* 4: 1-6.

Buchanan, J. M. and Tullock, G. (1975). "Polluter's Profits and Political Response: Direct Controls versus Taxes." *American Economic Review* 65: 139-147.

Buchanan, J. M., Rowley, C. K., and Tollison, R. D. (eds.) (1987). *Deficits.* Oxford: Basil Blackwell.

Buchanan, J. M. and Roback, J. (1987). "The Incidence and Effects of Public Debt in the Absence of Fiscal Illusion." *Public Finance Quarterly* 15: 5-25.

Buchanan, J. M. and Congleton R. (1998). *Politics by Principle Not Interest: Towards Nondiscriminatory Democracy.* Cambridge: Cambridge University Press.

Cohen, G. (1990) "Equality of What? Welfare, Goods and Capabilities." *Recherches Economiques de Louvain* 56: 337-382.

De Jasay, A. (1996). "Before Resorting to Politics." In *The Political Economy of the Minimal State.* Aldershot: Edward Elgar.

Downs, A. (1956). *An Economic Theory of Democracy.* New York: Harper and Row.

Dworkin, R. (198la). "What Is Equality? Equality of Welfare." *Philosophy and Public Affairs* 10: 185-246.

Dworkin, R. (1981 b). "What Is Equality? Equality of Resources." *Philosophy and Public Affairs* 10: 283-345.

Eusepi, G. (1989). "Buchanan's Critique of the Neoclassical Model and the Rediscovery of the Italian Tradition in Public Finance." *Rivista Internazionale di Scienze Economiche e Commerciali* 36: 801-822.

Eusepi, G. (1993). "Individualismo metodologico, mercato costituzionale e ' la via breve' della costituzione senza costituenti." *Rivista Internazionale di Scienze Economiche e Commerciali* 40: 641-656.

Fleurbaey, M. (1995). "Equal Opportunity or Equal Social Outcome?" *Economics and Philosophy* 11: 35-55.

Frey, B. S. (1997). *Not Just for the Money. An Economic Theory of Personal Motivation.* Aldershot: Edward Elgar.

Griziotti, B. (1917). "Diversa pressione del prestito e dell'imposta." *Giornale degli Economisti,* May. Reprinted in *Studi di Scienza delle Finanze e Diritto Finanziario,* pp. 198-261, Milan, 1956.

Kirchgaessner, G. (1994). "Constitutional Economics and its Relevance for the Evolution of Rules." *Kyklos* 47: 321-339.

Lindahl (1919). *Die Gerechtigkeit der Besteuerung. Eine Analyse der Steuerprinzipen auf Grundlage der Grentznutzentheorie.* Lund: Gleerups.

Mazzola, U. (1958). "The Formation of the Prices of Public Goods." In *Classics in the Theory of Public Finance,* Musgrave, R. A. and Peacock, A. T. London: Macmillan, 1958.

Meltzer, A. and Richard, S. (1981). "A Rational Theory of the Size of Government." *Journal of Political Economy* 89: 914-927.

Musgrave, R. A. (1938). "The Voluntary Exchange Theory of Public Finance." *Quarterly Journal of Economics* 53: 213-237.

O'Driscoll, G. P. (1977). ''The Ricardian Nonequivalence Theorem." *Journal of Political Economy* 85:207-210.

Pincus, J. (1977). *Pressure Groups and Politics in Antebellum Tariffs.* New York: Columbia University Press.

Puviani, A. (1903). *Teoria della illusione finanziaria.* Palermo: Sandron.

Racheter, D. P. and Wagner, R. E. (1999). *Limiting Leviathan.* Aldershot: Edward Elgar.

Sen, A. K. (1992). *Inequality Examined.* Cambridge: Harvard University Press.

Tollison, R. D. and Wagner, R. E. (1987). "Balanced Budgets and Beyond." In *Deficits,* Buchanan, J. M., Rowley, C. K., and Tollison, R. D., eds. Oxford: Basil Blackwell.

Wagner, R. E. (1987). "Liability Rules, Fiscal Institutions and the Debt." In *Deficits,* Buchanan, J. M., Rowley, C. K., and Tollison, R. D., eds., pp. 199-216. Oxford: Basil Blackwell.

Wicksell, K. (1958). "A New Principle of Just Taxation." In *Classics in the Theory of Public Finance,* Musgrave, R. A. and Peacock, A. T. London: Macmillan, 1958.

Chapter 4

GROWTH IN THE REAL SIZE OF GOVERNMENT SINCE 1970*

Thomas E. Borcherding
Claremont Graduate University
thomas.borcherding@cgu.edu

J. Stephen Ferris
Carleton University
stephen_ferris@carleton.ca

Andrea Garzoni
Carleton University
Agarzoni@ccs.carleton.ca

Abstract   This paper describes and comments on the growth of government since
                  the late 19th century, and describes and assesses various efforts to ac-
                  count for this growth. In keeping with the intent of this volume, the
                  focus is on the OECD nations generally, and not just on the U.S.

Keywords: Baumol effect, government size, growth of government

JEL classification: H11

1.   INTRODUCTION

In this chapter we survey what has happened to the real size of government
since 1970. We begin by simply describing what has happened to government
size in a sample of twenty countries from the Organization for Economic Co-
operation and Development (OECD) for the period following 1970. We follow
this with a selective summary of the current literature, especially that which
emphasizes the new factors and techniques used to explain the different pattern

---

of change that has arisen over this time period. In this regard, the traditional literature on government size assumes that causality runs from income to government size.

The second part of our work updates the estimates of the key parameters described in our literature review for a panel of twenty OECD countries. We begin by re-estimating the parameters of the demand curve for government services for the 1970–1997 time period. This allows us to inquire whether the changing pattern of government growth observed over this period has meant a break in the structure of the model determining government size. Next we reverse the direction of causality to examine the role of government size in relation to growth by estimating its impact within a simple growth model. While both government size and economic growth relationships have received considerable interest in their own right, less attention has been given to their interdependence. Without such recognition the two effects become co-mingled in each single equation coefficient. To make such a separation, we estimate the two equations simultaneously. This allows us to ask whether ignoring the simultaneity of this two-way relationship seriously biases the measure of the income effect (in determining government size) and/or the effect of government size on economic growth when each are estimated separately.

In the first two parts of this chapter we use the conventional measure of government size, i.e., aggregate government consumption from the Gross Domestic Product (GDP) side of the National Accounts divided by GDP, as our measure of the government's influence on the economy. In the final part of our survey, we consider whether this ratio has remained an appropriate index for the scale of government's activities. Most economists recognize that government has considerably more influence on society than the level of expenditure alone would suggest. In most countries, for example, governments set up and maintain (through legislation and often subsidy) operations which would otherwise be undertaken by private corporations (e.g., postal services, public utilities, hospitals). These frequently operate outside of the government's budget. In other cases, the government extends its influence by its granting (or withholding) preferential loans, import/export licenses etc. Finally, through its regulatory powers, governments can exercise a strong presence in the operation of the economy without its role ever appearing as an expenditure item. In this sense the observation that the traditional measure of government size is leveling off or even shrinking may well mislead if the role of government has simply changed from one of direct spending to one of indirect influence through regulation.

2.   RECENT CHANGES IN THE SIZE OF GOVERNMENT

We begin our historical survey of government size by noting that unlike earlier periods, it is no longer the case that real government size has only grown. While the average annual rate of growth for our group of twenty OECD countries has remained marginally positive for the period since 1970, individual countries in our sample have experienced widely different growth rates.[1] In the United States, for example, the share of government consumption in GDP fell—from 18.5% in 1970 and to roughly 15.5% by 1997. In the period following 1980, Belgium, Italy, and the Netherlands experienced similar declines. As the positive average growth rate does imply, however, a fall in the size of government was neither universal nor even typical. Particularly since 1975, government size has grown rapidly in such OECD countries as Austria, Finland, France, Greece, Ireland, Norway, Portugal, Spain, and Switzerland. Countries that experienced no overall change in size were least common, with Sweden and Korea representing these special cases.[2] Finally, for at least some subset of countries, real government size has appeared to peak. Australia, Belgium, Canada, Germany, Italy, Japan, the Netherlands and U.K. all experienced an initial period of growth followed either by no change or by a period of contraction. With such a variety of different outcomes straightforward generalization becomes problematic. Perhaps the safest generalization is simply that the pattern of growth in government size since 1970 has been much more varied than the pattern of continuous growth experienced in the period prior.

2.1.   The Literature on Real Government Size

We take as the starting point for our survey of the more recent literature, Borcherding's survey articles (1977 and 1985) on the determinants of government size. To explain the continuous rise in U.S. government size through 1970, Borcherding derived the following equation:

$$\dot{g} = (\eta + 1)\dot{p} + (\delta - 1)\dot{y} + (\alpha\eta + \alpha - 1)\dot{N} + \eta\dot{t} + \delta\dot{k} + \phi\dot{m}, \qquad (1)$$

where the dots above the variables signify rates of growth and where g is the share of government spending in aggregate real output, p is the relative price of government services (to all other goods), y is mean income, N is population size, t is the share of the cost of government borne by the median voter, k is the ratio of median to mean income, and m is a set of political control variables. The parameter η represents the price elasticity of demand for government consumption; α, the degree of publicness of the output of the government sector; δ, the income elasticity of demand and; φ, the set of elasticities

for the effect of the various political controls on demand. If the median voter pays a representative share of the cost of government, then (1) simplifies to

$$\dot{g} = (\eta + 1)\dot{p} + (\delta - 1)\dot{y} + (\alpha - 1)(\eta + 1)\dot{N} + \delta\dot{k} + \phi\dot{m}, \tag{2}$$

a useful form for estimating the effects of the different variables on real government size. In addition, equation (2) provides a convenient taxonomy for discussing the research done on real government size and the consensus that has grown up over the size of the model's key parameters. The time period that follows Borcherding's work can be characterized as one of moving from single equation to system estimation with greater emphasis being placed on time series issues and wider use of panel data. Having said this, the parameters of greatest interest to most public finance economists remain those emphasized by Borcherding. For this reason, then, we organize our survey of the recent literature in terms of these parameters.

2.2.   The Elasticity of Demand for Government Services and the Baumol Effect

Discussion of the relationship between real government size and the price of government services is now an integral part of the debate over Baumol's Cost Disease hypothesis.[3] Baumol (1967) hypothesized that because the output of the government sector is relatively labor intensive, its rate of productivity growth would be expected to be low relative to that of private sector output. This implies that over time the real cost of public sector output will rise relative to all other goods. It then follows that if the demand curve is price inelastic, a rise in the relative price of government services will result in only a relatively small decrease in the quantity of government services demanded and hence a higher aggregate expenditure on public sector output. Studies by Bradford, Malt and Oates (1969), Beck (1979), Spann (1977), Peltzman (1980), Berry and Lowery (1984), Ferris and West (1996b, 1999) all have documented the steady rise of the relative cost of U.S. government services. Borcherding (1977) calculates the U.S. pre-1970 average growth rate to be about 1.5% per annum and our calculations from a more recent panel of data through 1997 (presented in the next section) suggest that while the growth trend has fallen, it has remained positive.[4] With positive growth in the relative price of government services through time, a necessary condition for the emergence of the Baumol effect is that the parameter estimate on $\dot{p}((\eta + 1)$ from equations (1) and (2)) be between 0 and 1.[5] Early estimates of the price elasticity by Borcherding and Deacon (1972) and Bergstrom and Goodman (1973) point to a value of $\eta + 1$ of around 0.5 (or a value of $\eta = -0.5$). Borcherding's (1985) summary of U.S. federal, state and local expenditure data from 1902 to

1979 found the average value for η closer to -0.4. Given this latter value for the elasticity parameter, Borcherding argued that if the government's share of GDP evolved as in equation (2), then the price effect alone would account for 31% of the growth in U.S. government size that took place between 1902 and 1979.

While many have debated the reasons why the real cost of government services has risen in the past and Baumol's accompanying prediction that this will continue into the future,[6] the hypothesis that the demand curve for government services is price inelastic is now standard in the literature.[7] One might have thought, however, that the typical finding of a low value for the price elasticity of demand in a single equation model could well have arisen from the inability of a single equation technique to separate out offsetting demand and supply influences on observed outcomes. Simultaneous re-estimation of the demand and supply system, however, only confirms what is implicit in the single equation approach, namely that variations in a constant cost supply curve trace out positions of equilibrium along a stable demand curve.[8]

2.3. The Income Effect and Wagner's Law

One of the oldest ways of explaining public sector growth is associated with the well known German economist Adolph Wagner [1835–1917] and what is commonly known as Wagner's Law, or, the "law of expanding state expenditure." Wagner noticed "empirical regularities" in the growth of central, local and public enterprises expenditures and observed there appeared not only to be an absolute but also a relative expansion of the public sector as economies develop.[9] Wagner's ideas have motivated a large number of studies in the literature. This section reviews a small part of this literature, with greater emphasis given to newer studies that pay greater attention to the time series problems in the actual data.

Confusion sometimes surrounds the testing of Wagner's Law because different authors use different specifications of the test. Some of the earliest studies, in particular those by Musgrave (1969) and Goffman and Mahar (1971), test for the presence of Wagner's Law by looking at the ratio of government spending relative to per capita income, our $\delta - 1$ in equation (2) above. In this form of the test, their finding that this elasticity was greater than zero was interpreted as yielding support for Wagner's Law. Gupta (1967) tested real government expenditure relative to real income for five different countries (U.S., U.K., Sweden, Canada, Germany) and finds an income elasticity, the equivalent of our δ, to be greater than unity. Bird (1971) used a similar specification for four countries (U.K., Germany, Sweden and Japan) over different subperiods and found evidence supporting Wagner's Law. He estimates income elasticities ranging from 1.02 for Japan to 3.90 for Germany in his most

recent subperiod. Ganti and Kolluri (1979) formulate their test in per capita terms in relation to the U.S. and find a δ of around 2.[10]

Gandhi (1971) looks at cross-section studies of Wagner's Law, and finds that Wagner's Law appears to hold for a sample of both rich and poor countries, but does not hold if only the poorer countries are taken into consideration. Ganti and Kolluri (1979) argue that Wagner's Law requires not only an income elasticity greater than unity but also a rise in the per capita quantity and/or quality of public services. The latter condition is not met in a sample of only less-developed countries.[11] Gandi's finding seems to accord to what Abizadeh and Gray (1985) find when they test the hypothesis for a pooled time series/cross section sample of 55 countries. By dividing the sample in three groups according to level of GDP per capita, they find support of Wagner's Law for the two richer groups but not for the poorer group.

Finally, while most studies have looked at income elasticity relative to aggregate public expenditure, Borcherding and Deacon (1972) test equation (2) above on U.S. data at the state level for a wide variety of different public goods and services. They find income coefficients, i.e., values for $\delta - 1$, that range from 0.0421 for sanitation services to 2.7359 for parks and recreation. Since the elasticities for all seven expenditure groups were greater then zero, this was interpreted as support for Wagner's Law.[12]

In evaluating these studies, one must be aware of potential methodological problems. From an econometric point of view, we must take into account that when time series data is used the underlying variables are often not stationary in levels. In the case of Borcherding and Deacon (1972) this is not a problem, since their study looks at data at a particular point in time, namely 1962. However, when the study's purpose is to analyze the evolution of the government's share of GDP over time, then the time series properties of the dependent as well as the independent variables must be recognized in order to make correct statistical inferences on the estimated parameters.

These considerations have spurred a new wave of studies on Wagner's Law where more explicit attention is given to the time series properties of the data. From this perspective, one of the major shortcomings of the older literature on government growth has been the implicit assumption that the respective time series were stationary in their levels. Often this is not true, following stochastic processes that contain unit roots. In such cases, ordinary least squares (OLS) estimations done on the level values of these variables yield inconsistent estimates of the income elasticity if the two series are not cointegrated. In addition, such spurious regressions tend to be characterized by artificially high values of the R^2 and low Durbin-Watson statistics. On the other hand, if the two time series are cointegrated, then the problem of spurious regression in the sense of Granger–Newbold does not arise.[13]

It follows that one way of approaching the data is to run an Augmented Dickey-Fuller (ADF) test on the levels of the variables to check for the presence of a unit root. If that test indicates that the null hypothesis of a unit root cannot be rejected at a reasonable confidence level, we may conclude that the series are non-stationary in levels. If these variables allow rejection of the null when run in first differences, the variables are stationary in first differences or integrated of order one, I(1).

If the individual series are nonstationary in levels, we can proceed by testing whether the series are jointly cointegrated. This is done by regressing one series on another and applying the ADF test to the residuals. If the ADF result allows rejection of the null of a unit root in the estimated residuals, then we can say that the two series are cointegrated of order one or CI(1, 1). Under these conditions, an error correction model can be formulated and Wagner's Law may be tested through Granger-causality tests. More specifically, if we find that per capita income Granger-causes government size, then this is evidence that Wagner's Law holds.

Among the new studies that have approached Wagner's Law in this manner are Henrekson (1993), Bohl (1996) and Payne and Ewing (1996). Henrekson (1993) tests for Wagner's Law in Sweden using data from 1861 to 1990. He finds that that the levels of the two variables, real government size and per capita income, are not stationary but become so upon first differencing. In addition, Henrekson finds that the two series are not cointegrated, so that no consistent estimate of income elasticity can be estimated. He concludes that no long-run relationship can be established for Sweden and hence that it is unlikely that "growth in real income *per se* caused the growth of government."[14]

Bohl (1996) tests for evidence of Wagner's Law on G7 countries using primarily post-World War II data.[15] He finds that all the time series variables are I(1). Furthermore, he finds evidence of a long-run relationship only for Canada and the U.K. In all the other countries, the null hypothesis of non-cointegration cannot be rejected. Bohl then proceeds to test for Granger causality in these two countries alone and concludes that that since real per capita income Granger-causes government size, Wagner's Law is supported.[16] Payne and Ewing (1996) use an error correction model to test for Wagner's hypothesis on a sample of 22 randomly selected countries. Evidence of Wagner's Law is found only for Australia, Colombia, Germany, Malaysia, Pakistan and the Philippines. Bi-directional causality is found for India, Peru, Sweden, Switzerland, U.K., U.S. and Venezuela and Granger causality is absent in Chile, Finland, Greece, Honduras, Italy and Japan.[17]

2.4. Income Inequality between Mean and Median Voter

Another explanation of the growth of the public sector utilizes the political/electoral factors that underlie public choice theory. In this framework the

size of government is determined, in part, by the rules and procedures of the voting process that lead to the resolution of political choices and where the allocative outcome of that voting process is strongly influenced by the distribution of income. The intuition for including income inequality as a determinant of government size has been developed in detail by Meltzer and Richard (1981). To do so they assume that government sector goods and services are pure public goods and serve a purely redistributive function. Then, under a majority-voting rule, the decisive voter in determining the scale of government becomes the median income earner.[18] The median voter sets the tax share and hence the amount of redistribution. Various studies on income distributions have confirmed that the typical distribution is skewed to right, so that median income is typically below the mean income.[19] The consequences for government size in this model are that voters with income below the median will always favor increased redistribution so that any change that increases mean voter income relative to median voter income will also increase government expenditure. Meltzer and Richard (1981) argue that changes increasing government size include such structural and demographic changes as the extension of suffrage to lower income families earlier in the last century and more recent increases in the proportion of retired voters and their effect on size through the social security system.[20]

If this were the only dynamic at work, redistribution would stop only when the median voter succeeded in redistributing enough to become the mean voter. There is another factor at work, however that does constrain the amount of redistribution. This is the disincentive that higher taxes create on the incentive to work and hence on the income available to be redistributed. Higher taxes then become the mechanism that effectively limits the scale of redistribution.

In a follow-up article, Meltzer and Richard (1981) test their model on U.S. data and find a positive relationship between the level of government expenditures and both the level of median income and the ratio of mean to median income. Later empirical research, however, has yielded inconsistent findings. Henrekson (1988) and Lybeck (1986) both reject the Meltzer–Richard hypothesis. On the other hand, Henrekson (1990) and Henrekson and Lybeck (1988) find support for the role of income distribution as a determinant of government growth for Sweden. In pooled cross-sectional, time-series data Kristov, Lindert and McClelland (1992) even find a negative coefficient for the income distribution variable, and in cross-country context Mueller and Murrell (1985, 1986) find no more than weak evidence in support of the Meltzer–Richard model.

For French time series data, Aubin et al. (1988) do find evidence supporting the role of the rent-seeking hypothesis as put forth by Meltzer and Richard.

In this context it is worth mentioning the work of Peltzman (1980). Peltzman models the growth of government as the result of changes in between-group and within-group income inequality. Either an increase in the former or

a decrease in the latter could explain growth in government size. Empirically, Peltzman finds that the latter factor has been most prominent. In particular, an increase in within-group equality, i.e., the growth of a homogeneous "middle class," has played a dominant role in explaining the growth of the public sector following World War II.

2.5. Other Factors Influencing Either the Demand and/or Supply of Government

Much recent work on the determinants of government size has been devoted to analyzing the significance of the final set of political control variables, the m's in equation (2).[21] We highlight two particular approaches. First, increasingly attention is being given to the incorporation of electoral politics and the role of interest groups into both formal models of public choice and their empirical counterparts.[22] In practice, many authors now incorporate political variables to control for the effects produced by changes in the strength of political interest groups who have an incentive to alter the real size of government.[23] These would include segments of the population that typically benefit from larger government size, such as the fraction of the population who are poor, disadvantaged and/or older. Others would include more organized groups who expect to benefit (lose) from a further expansion in the role of government, such as farm or urban lobbies, union groups etc. Finally, as government size has grown, so has its work force. There is then a direct incentive for government employees to vote for larger government and this feedback has also been incorporated into the analysis of government size.[24]

A second approach to finding relevant control variables has emphasized changes in the relative cost of raising funds. This is the particular focus of work by Kau and Rubin (1981). Their approach would suggest that such factors as a rise in the participation rate by women and the movement of economic activity from the farm to the city have lowered the cost to government of raising funds while the rise in self employment has raised the cost of collecting funds. To this list, and in anticipation of the work in later sections, we add the degree of openness of an economy as a constraint on the ability of government to raise tax revenues and hence as a constraint on government size. All of these factors have been found to be significant in their effect on government size.[25]

2.6. Government Size within a Growth Equation

While our attention has been focused on the determinants of government size, much of the recent work on government size has reversed the direction of causality to investigate the role of government size in relation to economic growth. This is in part a response to the growing interest in the empirical determinants of economic growth and in part because of the greater use of pure

time series techniques (see Section 2.4 above).[26] From the perspective of the growth literature, however, the typical finding is that larger government size (measured particularly in terms of government consumption) lowers economic growth.[27] The importance of this finding for our work is that it suggests that the income coefficient in a single equation model of government size may incorporate too much of the causality running from size to income. The coexistence of competing theories of causality suggests that the two should be estimated simultaneously in order to determine the separate size and significant. We do this in the following section.

3.   EVIDENCE FROM PANEL DATA ON TWENTY OECD COUNTRIES FROM 1970 TO 1997

In this section we present the results of estimating a regression model of real government size along the lines of that outlined by Borcherding (1985) and presented as equation (2) above. Our first objective is to see if the consensus parameter values found for U.S. experience prior to 1970 bear any resemblance to those found for a panel of OECD countries over the post 1970 time period. In essence we are asking whether the variety of cultural, political, and specific time-period effects experienced across this set of OECD countries has resulted in a variety of growth patterns that resists incorporation into a single underlying theory. Our second objective is to ask whether the fact that government size and economic growth are determined simultaneously means that a model that focuses on one-way causality cannot measure accurately the separate contributions of each determinant of government size or economic growth. To do so we estimate a simple growth model where government size is a key determinant of economic growth, hence inverting the causality assumed in the government size equation. Here we find a Solow-type growth model performs well as an explanation of real income and output growth. The third objective is to put the two relationships together by jointly estimating the two equations as part of a system under three stage least squares. This allows us to determine whether there is likely to be a significant bias when the simultaneity of the relationship is not accounted for econometrically.

3.1.   Determinants of the Real Size of Government

In Table 1 we present a series of single regression equations that test the model of government size outlined in equation (2) above.[28] All equations utilize White's adjustment for heteroskadisticity, present the standard errors of the coefficients in brackets below each coefficient estimate, and include the important elasticities of the underlying demand equation (implied by the coefficient estimates) at the bottom of each equation column. Our equations utilize

TABLE 1.
Panel Regressions on the Determinants of the Growth Rate in Government Size.
1970–1997 for Twenty OECD Countries Ordinary Least Squares (OLS) with White
Heteroskedasticity-Consistant Standard Errors. The dependent variable is the growth
rate of (g/y)

Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.010*	0.020*	0.022*	−0.010	0.353**	0.466**
	(0.002)	(0.003)	(0.004)	(0.006)	(0.154)	(0.227)
Growth(y/N)	−0.563*	−0.534*	−0.581*	−0.576*	−0.506*	−0.608*
	(0.047)	(0.040)	(0.047)	(0.041)	(0.056)	(0.063)
Growth(p_g/p)	1.023*	0.989*	0.918*	0.999*	0.983*	0.978*
	(0.059)	(0.061)	(0.070)	(0.045)	(0.063)	(0.062)
Growth(N)	0.146	−0.224	−0.082	−0.154	−0.343	−0.680
	(0.242)	(0.226)	(0.239)	(0.373)	(0.206)	(0.379)
D(Openness)		−0.097*	−0.098*	−0.105*	−0.059**	−0.082*
		(0.026)	(0.030)	(0.024)	(0.027)	(0.027)
Time		−0.001*	−0.001*	−0.001*	−0.024*	−0.034**
		(0.0002)	(0.0002)	(0.0001)	(0.010)	(0.0156)
D(Oldpop)		1.531**	1.413	0.429	1.741*	−0.016
		(0.688)	(0.848)	(0.708)	(0.667)	(0.846)
D(Self)		−0.231**	−0.219	−0.166	−0.238*	−0.196**
		(0.099)	(0.183)	(0.105)	(0.091)	(0.096)
Growth(Gini)			0.0354			
			(0.057)			
Fixed Effects: Countries	No	No	No	Yes	No	Yes
				$F = 4.47$		$F = 4.88$
				Prob = 0		Prob = 0
Fixed Effects: Time Period	No	No	No	No	Yes	Yes
					$F = 3.80$	$F = 4.50$
					Prob = 0	Prob = 0
No. of Observations	540	473	368	473	473	473
Adjusted R^2	0.542	0.573	0.583	0.623	0.629	0.678
Standard error of Regression	0.0255	0.0223	0.0218	0.021	0.021	0.019
Implied Elasticities:						
Price, η	0.023	−0.011	−0.082	−0.001	−0.017	−0.022
Income, δ	0.437	0.466	0.419	0.424	0.494	0.392
Publicness, α	1.146	0.774	0.911	.846	0.652	0.307

Statistically different from zero at 1% (*), 5% (**).
Source: OECD Statistical Compendium (CD Rom, 1999). National Accounts I and Labour Market and Social Issues: Annual Labour Force Statistics (see Data Appendix) Countries include Australia, Austria, Belgium, Canada, Finland, France, Germany, Greece, Italy, Ireland, Japan, Korea, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, U.K., U.S.A. Factor Income Gini Coefficients from Milanovic (2000, Appendix).

up to five explicit public choice control variables: the fraction of the population older than sixty five (Oldpop), the fraction of the labour force that is self employed (Self), the ratio of exports to GDP (Openness), the Gini coefficient (Gini), and finally time.[29] The first of these variables, Oldpop, is designed to capture the political demand for social services by the older proportion of the public and is expected to be positively related to real government size. Self is expected to capture an important relative cost of tax evasion, since greater self employment gives individuals a greater opportunity to hide income and/or expense consumption (Kau and Rubin, 1981). The ex ante Gini coefficient is used to proxy the departure of median from mean income and Openness controls for what Rodrik (1998) calls an important empirical regularity between a country's exposure to international trade and the size of its government.[30] Its inclusion is of interest in its own right because of the complicated way in which openness can influence both government size and economic growth individually and in combination.[31]

Equation (1) presents our benchmark model of the traditional determinants of the growth in real government size. Equations (2) and (3) add the set of public choice variables and the Gini coefficient. Equations (4) through (6) utilize the panel feature of the sample by allowing for country-specific and time-period specific fixed effects, first individually and then in combination. Not that even our most basic equation explains more than fifty percent of the variation in the growth of real government size and, as the different fixed effects are added to the analysis, the equations come to account for almost seventy percent. This can be compared to Borcherding's (1985) finding that such equations could explain roughly fifty percent of the variation in size. What is also impressive is the consistency of the model's key coefficient estimates— i.e., across the six different models all of the estimates of the income and price elasticities are quite similar. Finally the fixed effects are significant both individually and in combination. Thus despite the finding of significant differences across OECD countries and across time-periods, the core predicted relationships between per capita income, relative cost and government size remain consistent in their estimated effect.[32]

Not only are the elasticity estimates derived from our OECD panel relatively constant across equations but the values of these estimates are also remarkably similar to those reported earlier for the U.S. in the pre-1970 period. In particular, our income elasticity findings (see the estimates of δ in the second last row) are centered about Borcherding's (1985) finding that the average value of the income elasticity of demand, δ, was about 0.400 (and hence inconsistent with Wagner's Law prediction of an income elasticity greater than one).[33] Similarly the price elasticity of demand, η, is insignificantly different from zero for all our estimates and so consistent with most other studies

that have found the demand curve for government services to be highly in-elastic. Finally, our typical finding in relation to publicness, i.e., that $\alpha \approx 1$, is consistent with other work that has found little "publicness" in the nature of government goods.[34] However, it is of interest to note that when both country and time-period fixed effects are present (i.e., the results in column (6)), more evidence of publicness is indicated than has usually been found.

It is only in relation to the fourth term in equation (2), i.e., testing for the prediction arising from changes in the distance between the mean and me-dian voter, that we have been largely unsuccessful. Here the recent work of Milanovic (2000) on *ex ante* (factor income) measures of income distribution seemed to offer one potential test of the median income hypothesis. Greater income inequality (i.e., a larger Gini coefficient) indicates a larger discrep-ancy between median and mean income and hence, through the median voter model, a larger demand for government services and thus a larger real size of government. A time series of *ex ante* Gini coefficients would then allow us to estimate a coefficient for $\dot{k}$ in equation (2) above. Unfortunately, very few fac-tor income Gini coefficients are available (Milanovic finds a maximum of five for some countries, with most countries having two or fewer observations). By interpolating between observations for the countries in our sample we did construct a time series (called Growthgini in Table 1). However its inclusion in the equation of columns (3) adds little to the explanatory power of the model. We attribute this to the sparsity of data rather than the inappropriateness of the test.[35]

The set of public choice variables (in (2)) do contribute significantly to the explanatory power of the basic government size equation.[36] In terms of the individual predictions, Openness is always negative and significantly so in all equation estimates.[37] This is then consistent with the public choice prediction that greater openness will expose government to greater tax/service competi-tion and hence will constrain government size by raising the cost of collecting funds. The two other specific public choice variables, Oldpop and Self, are also significant determinants of government size before the different fixed effects are investigated. However while the fraction of the labor force that is self-employer typically has a significant negative effect on government size, the significantly positive effect of an aging population on size tends to disappear once the equations incorporate country specific dummies.

Finally, note that both the constant and an explicit time trend term remain significant throughout. This suggests the existence of some time-related but currently unexplained process that has significantly affected real government size. Holding constant the other coefficients in the equations, the two time co-efficients suggest an inverted-U shaped pattern for government size over the sample period. The estimates of Columns (2) and (3) suggest that government size would have peaked towards the end of our time period (between 1995 and

1997), while the equations incorporating fixed effects suggest a peak somewhere more in the middle (between 1982 and 1985).

What can we conclude from this exercise? First, without resorting to the wider range of public choice variables now used to supplement explanations of government size, even the smallest set of traditional economic variables performs well and can explain at least fifty percent of the variation in the growth of real government size over this period. The addition of a small subset of potential public choice variables has a significant but quantitatively small effect on the explanatory power of the equation. Second, the estimates of the underlying coefficients are not substantially different from those found earlier and hence are not specific to U.S. experience in the period prior to 1970. That is, the period following 1970 does not appear to represent a statistical break from earlier periods, despite the fact that the direction of change is no longer only upwards. Third, the results suggest that there are significant country and time period effects. However such a finding does not preclude us from confirming the existence of a common, consistent pattern to government growth across countries through time.

3.2.   The Effect of Public Sector Size on Growth

In Table 2 we present a series of panel regressions that focus on the relationship between real output growth and government size. These regressions allow for a constant across time periods and countries and, in the later regressions, allow this constant to vary by country and time-period. In these models the variable representing government size is its rate of change rather than its level. This is because government size is nonstationary for many of the countries in our sample and the use of its level (rather than first difference) would introduce the time series problems discussed earlier.[38] For the same reason we include the terms-of-trade variable in its first difference. However, it should also be pointed out that when the lagged value of the logarithm of real per capita income is included to test for convergence, the same time series issue is reintroduced. We are not sure how to deal with this problem.[39]

A simple Solow representation of the growth process suggests that real output growth will be driven by the rates of growth of the underlying factors of production and variations in the savings rate. For this reason our basic growth equation includes the ratio of gross capital formation relative to GDP (investment = savings rate) and the population growth rate. Most growth equations also test for conditional convergence and to incorporate this test we include the lagged value of real output (despite the time series problem that this introduces). While most such equations would incorporate a measure of human capital, our use of annual data precludes the use of the Barro-Lee measure (most often used for this purpose).[40] Finally we include an additional variable

TABLE 2.

Panel Regressions for Growth Effects 1970–1997 for Twenty OECD Countries. OLS with White Heteroskedasticity-Consistant Standard Errors in brackets. Dependent variable is the Growth in Real Output (i.e., $\dot{y}/y$)

Independent Variables	(1)	(2)	(3)	(4)
Constant	0.264*	0.332*	0.236*	0.219
	(0.032)	(0.061)	(0.034)	(0.116)
Growth(g/y)	−0.377*	−0.365*	−0.297*	−0.276*
	(0.028)	(0.038)	(0.028)	(0.035)
Lnrypc^{-1}	−0.028*	−0.034*	−0.025*	−0.020
	(0.003)	(0.006)	(0.003)	(0.013)
Growth(N)	0.454**	−0.066	0.373**	−0.204
	(0.199)	(0.381)	(0.168)	(0.380)
D(TermsofTrade)	0.065*	0.069**	−0.007	−0.003
	(0.023)	(0.025)	(0.021)	(0.166)
Savings Rate	0.135*	0.153*	0.133*	0.142*
	(0.025)	(0.042)	(0.021)	(0.048)
Fixed Effects: Countries	No	Yes	No	Yes
		$F = 1.34$		$F = 2.43$
		Prob(0) = 0.16		Prob(0) = 0
Fixed Effects: Time Periods	No	No	Yes	Yes
			$F = 10.51$	$F = 17.7$
			Prob(0) = 0	Prob(0) = 0
Regression Statistics:				
Adjusted R^2	0.404	0.414	0.588	0.609
Standard Error	0.023	0.023	0.019	0.019
Observations	540	540	540	540

Significantly different from zero at 1% (*), 5% (**).
Source: OECD Statistical Compendium (CD ROM, 1999). National Accounts I. Countries include Australia, Austria, Belgium, Canada, Finland, France, Germany, Greece, Italy, Ireland, Japan, Korea, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, U.K., U.S.A.

to capture the potential effects of external trade on growth. This is the change in the terms of trade (D(termsoftrade)). The hypothesis is that a positive trade shock should spur domestic production and hence growth. D(termsoftrade)) is then expected to be positively related to growth.

Before discussing the estimated coefficients, it is interesting to note that the addition of country-specific fixed effects have no significant impact on the equation's explanatory power, suggesting a degree of similarity in the growth process across countries that is not present in the analogous equation for government size. Specific time-period effects (common time effects across countries) do have significant explanatory power and, in combination, the two effects are found to be significantly different from zero.

In terms of the traditional variables in the growth model, the model without any fixed effect (Columns (1)) finds, as expected, that both the savings rate and the growth rate of labor are significantly positive determinants of output growth. This remains true when time-specific fixed effects are allowed for (as in Column (3)). However, when country specific fixed effects are added to the model (in Columns (2) and (4)), the population growth rate term loses both its sign and significance. The savings rate coefficient, on the other hand, remains significantly positive throughout, with the size of its coefficient tending to rise when the country dummies are present. Finally, all four equations are consistent with the conditional convergence hypothesis. The lagged value of the logarithm of real output is significantly negative in all versions of the growth model.[41]

In relation to external trade our results are mixed. The terms-of-trade effect is significantly positive (as expected) in the first two equations but the sign and significance change when the specific time-period fixed effects are introduced. Because the different time periods account for significant common effects across countries, the time period dummies may pick up the common component of the business cycle in our sample. To the extent this is what is being captured, the simultaneous fall in importance of the terms-of-trade variable suggests that at least part of the business cycle may be transmitted through external shocks to relative trade prices. At least in this form, however, the growth equation is not consistently supportive of the hypothesis that growth responds positively to favorable terms-of-trade effects.

The variable of primary interest to us is the effect of real government size on economic growth, where again our measure of size is government consumption expenditure relative to GDP. Holding constant the control variables discussed above, Table 2 indicates that an increase in the growth rate of real government (consumption) size has a consistently significant negative effect on real output growth. The size of that effect is reasonably constant across the different forms of the regression equation, diminishing only slightly from -0.377 to -0.276 as the fixed effects are introduced. The equation results are then consistent with the hypothesis that government consumption size is inimical to economic growth.

3.3.   Simultaneity in the Growth of Government Size and Real Income

While each model hypothesizes a negative relationship between the growth rate of government size and the growth rate of real output, the two models separately imply a direction of causality that runs opposite to the other. In this section, then, we re-estimate the two models as a system to determine the separate contribution of each part to the overall result.

Before presenting these finding, however, we consider the change that should be found in the size of the two coefficient estimates if the two-way causality suggested by each separate theory is jointly present in the data. First, if there is an important feedback from the growth of government size to income growth (as suggested by the growth equation), then we should find that the size of the income growth coefficient in the government size equation has been understated (in absolute terms) when estimated in isolation. That is, an exogenous increase in income growth will reduce the growth rate of government size, but that reduction in government size will produce a further increase in the growth of income. Then because the single equation attributes the change in government size to the full change in income, the inclusion of the reinforcing effect will understate the size of the uni-directional effect of income on government size. This in turn implies that the estimate of the income elasticity of the demand for government from the single equation (i.e., Wagner's Law) should be biased upwards (since coefficient estimate equals $\hat{\delta} - 1$).

When we turn to the growth equation and consider the coefficient on government size, the two-way effect again reinforces so that the single equation estimate will again understate (in absolute terms) the size of the one-way effect. Here the growth in real government size lowers the income growth rate and that, in turn, further lowers the growth rate in government size. The incorporation of both these effects in the single equation reduced form estimate will then attribute too little power to the negative effect of government size on income growth. The re-estimated coefficient should be larger.

In Table 3 we present the panel regression results when the two equations are estimated simultaneously under three stage least squares.[42] Note that when the explanatory power of the government size and growth equations in Table 3 is compared to their single counterpart (in Tables 1 and 2), the government size model can be seen to retain more of its explanatory power than does the growth equation regression model.[43] This suggests that the government size equation may be more accurately specified than is the growth equation. Our findings for both equations, however, are not strongly in support of two-way causality, at least in the short run. The growth equation, in particular, shows little evidence of the expected feedback. The final government size coefficient in Table 2 (0.276), for example, is of approximately the same size as that estimated in equation (3) of Table 3 (0.234) while being larger (rather than smaller) than the other three cases. Our findings after re-estimating the government size equation are more promising. Even though the coefficients of the simple form of the government size equation (in columns (1) and (2)) show little change when estimated simultaneously, the introduction of the country-specific fixed effects (and oil-shock dummies) do produce the expected effects on the size of the coefficient estimates for income growth. In equations (3) and (4) of Table 3, the elasticity estimates (-0.605 and -0.740) are both larger in absolute terms

TABLE 3.
Joint Estimation of the Growth of Government Size And the Growth Rate of Real
Income 1970–1997 for Twenty OECD Countries

Government Size Equation Dependent variable–growth of (g/y)	(1) TSLQ	(2) TSLQ	(3) TSLQ	(4) TSLQ
Constant	0.014*	0.013*	0.058**	0.028*
	(0.004)	(0.004)	(0.025)	(0.010)
Growth(y/n)	−0.305*	−0.265*	−0.605*	−0.740*
	(0.080)	(0.074)	(0.112)	(0.088)
Growth(p_g/p)	1.009*	1.030*	0.974*	0.950*
	(0.055)	(0.056)	(0.056)	(0.053)
Growth(N)	−0.102	−0.100	−0.183	−0.279
	(0.236)	(0.243)	(0.379)	(0.372)
D(Openness)	−0.099*	−0.095*	−0.103*	−0.108*
	(0.026)	(0.024)	(0.022)	(0.022)
Time	−0.001*	−0.001*	−0.001*	−0.001*
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
D(Oldpop)	1.282**	1.258**	0.441	0.460
	(0.626)	(0.612)	(0.647)	(0.630)
D(Self)	−0.158	−0.147	−0.046**	−0.110
	(0.109)	(0.108)	(0.024)	(0.100)
Fixed Effects:	None	None	Country*	Country*
Size Equation Statistics				
Adj R^2	0.536	0.521	0.610	0.610
S.E of Regression	0.023	0.023	0.021	0.021
η	0.009	0.030	−0.026	−0.050
δ	0.695	0.735	0.395	0.260
α	0.899	0.903	0.813	0.707
Growth: Dependent Variable is real income growth (y)				
Constant	0.258*	0.261*	0.266*	0.401*
	(0.060)	(0.046)	(0.045)	(0.084)
Growth(g/y)	−0.123**	−0.118**	−0.234*	−0.146*
	(0.053)	(0.053)	(0.047)	(0.051)
Lnrypc^{-1}	−0.027*	−0.027*	−0.028*	−0.039*
	(0.004)	(0.004)	(0.004)	(0.008)
Growth(N)	0.606**	0.602**	0.634*	0.045
	(0.261)	(0.259)	(0.248)	(0.404)
D(TermsofTrade)	0.095*	0.069*	0.034	0.036
	(0.026)	(0.026)	(0.025)	(0.025)
Savings Rate	0.114*	0.130*	0.115*	0.016
	(0.029)	(0.028)	(0.027)	(0.045)
Fixed Effects:	None	1974* 1979*	1974* 1979*	Country Effect** 1974*, 1979*
Growth Equation Statistics:				
Adj R^2	0.250	0.267	0.334	0.321
S. E. of Regression	0.024	0.024	0.023	0.023

Statistically different from zero at 1% (*), 5% (**).
Source: OECD Statistical Compendium (CD Rom, 1999). National Accounts I. Countries in-
clude Australia, Austria, Belgium, Canada, Finland, France, Germany, Greece, Italy, Ireland,
Japan, Korea, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, U.K., U.S.A.

than the corresponding coefficient in equation (4) of Table 1 (0.576). There is then some support for the hypothesis that the single equation estimate is likely to attribute too little explanatory power to the independent variable, at least in the case of the government size estimates. However such support is limited and even with this readjustment, the downward revision in the estimated value of the income elasticity of demand for government services only reinforces our earlier finding that its income elasticity is well below one, implying that Wagner's Law even less likely to be operative.[44]

Overall, then, the simultaneous equation estimates give little reason to believe that the usual single equation estimate will significantly understate the negative effect of government (consumption) size on income growth or the effect of per capita income on government size. In terms of contemporary causality, there is some suggestion of interdependence only for the case of government size. It remains unanswered whether more evidence of two way causality may reappear through dynamic interactions over time.

Turning to the other coefficients in the government size equation, three stage least squares estimation has made no discernable difference to the estimates of the size of the price elasticity of demand (compared to the single equation formulation). The price elasticity estimates remain close to zero, continuing to indicate a demand curve that is extremely inelastic. The system estimates suggest slightly more publicness in the nature of government consumption than did the single equation estimates (with the exception of of the final column in Table 1), however this tendency is strictly marginal. The implied values of the α coefficient never fall very far below one. In terms of the public choice variables, openness remains consistently negative in its effect on government size while both Oldpop and Self have their predicted sign but are only occasionally significantly different from zero. The cautioning note is again that the time coefficient remains a significantly negative determinant of government size, continuing to indicate the importance of some as yet unexplained time trend to the growth rate of government size.

When the coefficients other than government size in the growth regression are compared to their counterparts, only marginal differences are found. Simultaneous estimation has made no difference to our conditional convergence findings. Similarly the terms of trade variable retains its size and significance in the equations without fixed effects but, as in the single equation case, does loose its significance when country specific dummies are included. There is more of a change with respect to the estimates of the effect of labor and capital on growth-the coefficients on labor growth are now all slightly higher while their counterparts on capital (savings) are slightly lower. In addition, the labor coefficient repeats the pattern found in the early single equation regressions where the coefficient falls dramatically when country dummies allow for a different pattern of growth.

4.   "QUIET SIDE" OF THE PUBLIC SECTOR: REGULATION AS SPENDING

Posner in his classic 1971 paper "Regulation as Taxation" argues that fiscal instruments are but one of two instruments for executing public policy, the other instrument being regulation. Using Posner's methodology, a truer measure of public sector size would add to the budgetary costs, B, an estimate of the spending equivalent necessary to obtain private sector compliance with public rules and directives, R.[45] More generally, then, the real size of the public sector would be $g' = (B + R)/Y$, where Y is the measure of national product.

Leonard (1986) has also argued that public budgets will understate the size of the public sector by not recording what he calls the "quiet side" of public sector activity. He points to several sources of budget under-statement: promises of retirement benefits and social insurance, tax expenditures, subsidies in sales of public activities to favored groups but not others, and the aforementioned regulatory costs of government. He estimates that if one were to measure government at its full economic rather than budgetary cost, the U.S. federal government would be half again as big as its budget indicates! Somewhat earlier DeMuth (1980) had suggested similarly large figures. Unfortunately, the data behind these figures are rather more speculative than hard, so accepting them involves more an act of faith than an acceptance of evidence. For all writers, however, there are significant additional costs that the standard account of government size ignores.

Our task does not require us to assess the absolute size of government, g', only its change. Even this is not easy and our research for this survey has failed to find a comprehensive study of the regulatory or "quiet side" costs of this Posnerian measure. On the other hand, we have discovered three interesting measures of the regulatory impact of the U.S. federal government over the 1970–2000 time period. None is definitive or even fully comprehensive and so their use does not provide the precise measure for adjusting the more orthodox measure of $g = B/Y$ we are seeking. Nonetheless, we discuss their work in the hope that their presentation will shed some light on the issue and generate interest in further exploration by others.

While it has become common to observe that the number of U.S. federal government regulations has grown enormously relative to almost every measure of output, it does not seem sensible to assume that R is simply proportionate to the absolute number of rules and edicts.[46] It might not be too far from the truth, however, to assume that the size of the bureaucracy, as well as the size of the budgets of those bureaus needed to enforce these regulations, would in some rough-and-ready way be related to the full Posnerian cost of these regulatory agencies.

   With this in mind, we offer the evidence provided by, first, the Center for
the Study of American Business (CSAB) at Washington University (St. Louis)
and, second, studies coming out of the U.S. Office of Management and Bud-
get. The latest CSAB study (Warren and Weidenbaum, 1999) concentrates on
absolute measures of regulatory bureau expenditures and bureau employment
since the sixties and shows them to be quite large. However, when compared
either to the figures for non-military government spending or employment
(*Statistical Abstracts of the U.S.,* 1998), the relative change over our period
seems rather negligible. For example, in 1970 regulatory bureau spending was
1.2% of non-military federal spending. This has barely changed, rising to 1.3%
in 1980 but falling again to 1.2% in both 1990 and 1999. Relative employ-
ment measures, on the other hand, do suggest a significant increase in reg-
ulatory activity. The ratio of regulatory to non-military employment roughly
doubled, rising from 2.4% and 2.5% in 1970 and 1980 respectively, to 4.1% in
1990 and 4.7% in 1998. Unfortunately, no one has done the research to deter-
mine whether regulation at the state and local government level has changed
equivalently or whether the rising pattern of regulatory employment has been
matched in other OECD countries.

   Two other studies, one by Hopkins (1996) and a later update by Lutters
(1998), present different types of estimates of the impact of regulation on the
private sector and the costs of complying with U.S. federal regulation. Hop-
kins and Lutters both suggest large—in fact, huge costs of both regulatory
impact and compliance. However when measured *relative* to the economy as
a whole or relative to federal spending, they indicate almost no rise for the pe-
riod 1970 to 1998. Only Lutters can point to the Clinton years for a time when
his measure of regulatory cost rose relative to the economy as a whole.

   When we turn from studies that attempt to measure the full social cost of
regulation to studies that attempt the somewhat easier task of measuring the
degree of regulation across countries, several recent articles merit attention.
Cross country measures of the degree of governmental regulation, indices of
government effectiveness (Pryor, 2000), and the degree of regulatory burden
(Kaufmann, Kray and Zoldo-Lobaton, 2000) have been derived. For our pur-
poses, neither of these studies has been extended through time, nor is it clear
that their methodology could shed much light on the question of whether the
regulation has risen or fallen within any particular country. It is interesting,
however, to ask whether differences in the degree of regulation across coun-
tries can account for at least some part of the cross country differences in the
more traditional GDP measure of government size. More formally, we ask
whether the inclusion of a measure of the relative scale of regulation will im-
prove the explanatory power of a cross sectional equation of the determinants
of government size. Should it do so, its coefficient sign will indicate whether

TABLE 4.

Cross Country Regressions for Nineteen OECD Countries in 1997: Determinants of
Real Government Size and Real Government Spending per Capita. White
Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	t-Statistic	Coefficient	t-Statistic
C	0.407614	2.334007	5741.128	1.722076
RELPRICE97	−0.093959	−0.722773	−2111.882	−0.854333
RYPC97	−2.37E-06	−0.991257	0.119668	2.352735
POPULATION97	−2.52E-09	−0.025277	−0.000569	−0.246702
SWEDEN	0.341674	34.21050	6561.428	30.91194
REGULATION99	−0.054089	−2.383198	−1239.948	−2.461159
GOVTEFFECTIVE99	0.038362	2.966039	818.3873	2.990238
R-squared	0.926422		0.917582	
Adjusted R-squared	0.889632		0.876373	
S.E. of regression	0.029482		628.0657	
Sum squared resid	0.010430		4733599.	
Log likelihood	44.36153		−145.0045	
Durbin-Watson stat	2.134336		2.128318	

Dependent Variable: GSIZE; Dependent Variable: RGCPC; Method: Least Squares.
Sources: See Data Appendix.

regulation has been a substitute for, or complementary with, the more traditional spending measure of government size.

In Table 4 we present two cross-country regressions to investigate this question: one for the determinants of real government size, g/y; and the other for the determinants of real government consumption per capita. Both equations include the standard determinants of Section 3 plus Pryor's (2000) index of the relative degree of regulation across countries (REGULATION99) and his additional index of the relative effectiveness of government policies (GOVTEFFECT99). Because Pryor's measures are developed in 1999 and our most recent time period was 1997, our test assumes that the 1999 measures were relevant in 1997. Pryor's OECD countries also did not include Korea (and did include Denmark and New Zealand), so that these regression are for nineteen rather than the twenty OECD countries used earlier. Because the regressions are in levels, both equations also include a dummy variable for Sweden (whose government size is distinctly different from the other OECD countries).

What is most interesting about these two equations is that in both REGULATION99 and GOVTEFFECT99 are found to be highly significant determinants of government size and government per capita spending. The probability that neither variable adds to the explanatory power of the equation is only three percent.[47] Because Pryor's index of government regulation runs from 1 to 10

with high numbers corresponding to less regulation, the significantly negative coefficient signs for REGULATION99 in both equations means that that relatively more regulation is associated with larger government size. Regulation and government spending are complementary rather than substitutes for one another. Perhaps more straightforwardly, the significantly positive coefficient on GOVTEFFECT99 suggests that the more effective is government policy, the more government will be used and hence the larger will be government size.

Before leaving this topic it is also worth mentioning the literature that measures the growing size of the shadow economy and evaluates its connection with the hypothesis that the government's influence is becoming increasingly indirect, impacting increasingly through regulation. Friedrich Schneider, in a paper read to the Public Choice Society (2000), collected together the results of currency demand estimates of the relative size of the underground economy for 18 different OECD countries. In all of these countries he could show that the relative size of the underground economy grew, sometimes rapidly between 1970 and 1998.[48] In a more focused analysis of Austria, Schneider developed a methodology that could attribute the size of the underground economy to four underlying causes: direct taxes, indirect taxes, the complexity of taxes and the intensity of regulation. In the case of Austria, Schneider found that the percentage of the underground economy that could be attributed to regulation rose continuously from 11.8% in 1970 to 26% in 1995. To the extent that Austria is representative of other OECD countries and to the extent that the size of the underground economy measures the influence of government on the economy, Schneider's work reinforces earlier suggestions that more attention should be placed on regulation as a meter of government activity. It is likely that of the two types of Posnerian instruments, regulation has risen in relative importance.

To summarize our findings, the hypothesis that regulation-as-taxation estimates would change the measure of the relative growth of government is a proposition that cannot be readily sustained given the evidence at hand. Severe deficiencies of data and a dearth of research on the subject make the constant g' hypothesis since 1970 more speculative than authoritative. Although we are not confident enough to suggest that ignoring regulation-as-taxation does great harm for the last three decades of U.S. experience, we are more confident that ignoring the implicit tax effects of regulation over the longer 1900 to present interval will grossly understate both the relative size of government and its growth.[49]

5.   CONCLUSION

In this paper we have examined the real size of government in the period following 1970 and explored some of the literature developed to explain the

new growth patterns that have arisen. Our first important rinding is that the traditional measure of government size has no longer simply grown. A significant number of developed countries experienced either constancy or a reduction in their real government size. Despite this, we find that the earlier approaches taken to explain the growth in government over time are still robust in their ability to explain the newer variety of growth patterns that have emerged over our set of twenty OECD countries. In presenting these results we have also noted that in the modern growth literature, the consumption measure of government size has usually been thought to have a negative influence of income/output. Thus in the context of the government size literature, this suggests the presence of an important two-way relationship between government size and income that should be accounted for in the government size equation. Re-estimating the government size and growth equations as a system, however, indicates that the single equation estimates of the size of the income effect on government size and the effect of government size on growth do not change substantially. There is some evidence that single equation estimates may understate the effect of income on government size, but even the implied downward revision in the size of the income effect on government consumption does not change our conclusion with respect to Wagner's Law.

Finally our survey has explored the question of whether the GDP measure of government size may understate the "size" of the government sector due to the substitution of more indirect methods of control and influence for on-line budget expenditures. In effect, the "real" government size may still be growing if regulatory growth were incorporated properly in the measure. While data on this question is hard to come by and our current work suggests some reason for caution, our reading of the recent literature suggests that at a minimum the traditional measure of government size is becoming less useful as a measure of the government's influence over the economy. The importance of this question, combined with myriad of ways in which regulation both constrains and enhances economic performance, suggests that this a topic on which much more research is needed.

DATA APPENDIX

Data from the OECD Statistical Compendium (CD Rom, 1999),

i. National Accounts I.

gdppc   Gross Domestic Product per Head in US$—current prices and current Ppps
population   country population in hundreds of thousands

gcpc   General Government Consumption Expenditures per Head in US$—current prices and current Ppps.

gfcfpc   gross fixed capital formation per head in current prices and current Ppps

exchrate   national currency per US $

exportindex exports of goods and services ?? volume indices (1990 = 100)

gdpdefl   gross domestic product implicit price deflator (1990 = 100)

pg   government consumption price gdp price deflator (1990 = 100)

pi   gross fixed capital formation gdp price deflator (1990 = 100)

px   export (gdp) price deflator (1990 = 100)

pm   import (gdp) price deflator (1990 = 100)

gdpindex   volume index (1990 = 100)

ii Labour Market and Social Issues: Annual Labour Force Statistics

Civemp   Civilian labour force

Ownwork   Civilian employers and persons working on their own account

Pop65plus   Total Population aged sixty five years and older

Ceagric   Civilian employment in agriculature

Total Population LFSpop

All data is for 20 OECD countries from 1970 through 1997 (typically, 28 annual observations $*$ 20 countries = 560 observations). Some observations not available.

Countries: Australia, Austria, Belgium, Canada, Finland, France, Germany, Greece, Italy, Ireland, Japan, Korea, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, U.K., U.S.

Transformations of the Data   Most variables where transformed into natural logarithms, using the label Ln. Hence rates of growth were calculated as first differences in logarithms. Lngcpc = log(gcpc) = logarithm of government consumption per capita in current US$.

Both gcpc and gdppc were transformed into real variables by deflating by the <u>US</u>cpi. Eg.

Rypc = gdppc/uscpi and lnrypc = log(rypc)

Rgcpc = gcpc/uscpi and lnrgcpc = log(rgcpc)

Then, real size of government:

gsize = gcpc/gdppc and **lngsize** = log(gsize).

TermsofTrade = px/pm, **D(termsof trade)** = Termsoftrade – Termsof trade(–1).

Openness = exportindex/gdpindex, **D (Openness)** = Openness – Openness(–1).

Savings Rate = gfcfpc/gdppc.
Oldpop = Pop65plus/lfspop, **D(Oldpop)** = Oldpop − Oldpop(−1)
Self = Ownwork/Civemp, **D(Self)** = Self − Self(−1)
Other non OECD sourced data
Gini is taken from Milanovic B., 2000, *European Journal of Political Economy,* Appendix A. There the Gini coefficient is calculated on a per capita basis and the distribution is based on factor income prior to redistribution. Most countries had 3 to 4 observations and the Gini was interpolated linearly between observations and assumed to be constant both before and after the first and last observation. Of the twenty countries, observations were unavailable for Austria, Greece, Japan, Korea, Portugal, and Ireland and Switzerland had only one observation.
Regulation is the Total Index score of Table 1: Indices of Regulation-Laissez-faire for OECD Nations in the late 1990's. Pryor, 2000.
Government Effectiveness is the Total Index from Table A-2: Government Effectiveness Index, 1999. Pryor, 2000.

The numbers used after the variable names in Table 4 refer to the year. Hence Govteffect99 refers to 1999 value of the government effectiveness index and RYPC97 refers to real income per capita in 1997.

NOTES

1. The mean annual rate of growth of government consumption as a fraction of GDP was 0.682% for the twenty OECD countries in our sample (from 1971 through 1997).
2. Sweden continued to have (by far) the largest percentage of GDP in government consumption (53%) while Korea stood among those with the lowest (11%).
3. See Baumol (1967).
4. The average annual rate of growth of the price index of government sector output relative to the GDP deflator for the countries in OECD sample over 1970 and 1997 was 0.8 percent. See also Baumol (1993).
5. That is, $-1 < \eta < 0,$ implying that the demand curve must be inelastic.
6. See Beck (1979), Tyler (1996), Ferris and West (1996b, 1999), Tiongson (1997).
7. See Perkins (1977), Pommerehne and Schneider (1982), Gramlich (1985), Lybeck (1986, Ch. 5).
8. See, for example, Ferris and West (1996a).
9. See Peacock and Scott (2000), pp. 1-2.
10. See Henrekson (1993), p. 409.
11. See Gandhi (1971), pp. 53-55.
12. See Borcherding and Deacon (1972), p. 898.
13. See Granger and Newbold (1974), Payne and Ewing (1996, pp. 260-261), and Henrekson (1993, pp. 409-412).
14. See Henrekson (1993), pp. 412-413.
15. The G7 countries include: Canada, U.K., U.S., France, Italy, Japan, Germany.
16. See Bohl (1996), p. 196.

17. See Payne and Ewing (1996), p. 258 and p. 271.

18. For an alternative approach utilizing probabilistic voting theory see, Coughlin and Mankiw (1981) and Enelow and Hinich (1989).

19. A study particularly relevant to this article is "Income distribution in OECD countries" by A. B. Atkinson et al. (1995).

20. See Meltzer and Richard (1981), pp. 924-925.

21. Although not explicitly discussed in the text, political considerations are increasingly incorporated into growth analysis. See, for example, Zak and Knack (2001).

22. See, for example, Coughlin, Mueller and Murrell (1993) and Mueller (1993).

23. See the work of Buchanan and Tullock (1977).

24. Sometimes called the Buchanan/Tullock beaucratic-voting hypothesis. See also Ferris and West (1996a).

25. See Ferris and West (1996a). Their work emphasizes the importance of viewing these factors in combination rather than as alternative explanations of government size.

26. See, for example, Barro and Sala-I-Martin (1995).

27. See, for example, Kormendi and Meguire (1985), Karras (1993) and Folster and Henrekson, (1999, 2000). Empirical work that emphasizes the investment nature of governmental activity often finds a positive effect of size on growth. See Barro (1990).

28. Note that almost all these variables are nonstationary in their levels *so* that the use of either first differences and/or growth rates is the appropriate form in which to test the model.

29. Our use of government consumption rather than a more comprehensive transfer inclusive measure of government to measure size means that other traditional public choice interest group measures, such as the relative size of the farm population, will be less important as a determinant of size. In our equation estimates, relative farm population size was consistently positive in its effect but not significantly so.

30. See also Ades and Glaeser (1999), Weinhold and Rauch (1999) and Dinopoulos and Thompson (2000).

31. By including external conditions in the growth model below we hope to be able to distinguish between the effect of openness on trade and hence growth versus the effect of openness on government size and hence growth.

32. For fixed effects, the omitted country was the U.S.A. and the omitted year was 1970.

33. Peltzman (1980) points out that in relation to permanent rather than actual income, income elasticity is much higher.

34. Following Borcherding and Deacon (1972, p. 899), $\hat{\alpha} \cong 1 + [\hat{\theta}/(\hat{\eta}+1)]/[1+\hat{\sigma}_\eta^2/(\eta+1)^2]$.

35. Because of the greater use of proportional rather than majority voting in European countries (forming the large part of our sample), the median voter model's prediction may have inherently less explanatory power.

36. A Wald test of the hypothesis that the group adds no explanatory power can be rejected ($F=16.8$).

37. Note that in levels both openness (exports as a fraction of GDP) and government size (G/GDP) grow through time (see also Rodrik, 1998). However, because both are nonstationary in their level (and become stationary in first differences or growth rates) it is appropriate that openness enter the regression as a first difference.

38. That is, the level is I(1) while most of the other variables in the equation are I(0). The rate of change of real government size is I(0).

39. When the same regressions are rerun without the logarithm of lagged real income, all remaining coefficients increase slightly in absolute size while the adjusted R^2 falls from 0.404 to 0.310.

40. These are available only as five-year averages. The unavailability of an annual series for schooling may not be a serious cross-section problem since our countries are similarly developed, but it may present a bigger problem over time.

41. The equations were also run using the logarithm of the initial per capita income level (initialvalue) as a constant for each country. Such a procedure cannot separate a test of the convergence hypothesis from the use of country dummies to allow for country specific differences. One example, corresponding to Table 2 column (1) is:

Growthry = 0.180 − 0.357Growthgsize − 0.020initialvalue + 0.634Growthn + 0.067d(tot) + 0.135Savings rate

(0.027)(0.028) (0.0027) (0.201) (0.023) (0.026)

$\text{Adj}R^2 = 0.377$; SE = 0.024.

42. Table 3 presents only those combinations of county and time-period fixed effects across the two equations that permitted matrix inversion.

43. The adjusted R^2 of the government size equation remains largely unchanged while the adjusted R^2 of the growth equation falls considerably.

44. Again we note that the use of permanent rather than actual per capita income would affect the results in the opposite direction.

45. For example, if the state of California mandated through its Coastal Commission (CCC) certain land uses, the cost to private parties of carrying out that regulation—their compliance costs—would be the R. Another way of adjusting B would be to add an estimate of the budgetary costs of obtaining the easements and agreements to accomplish the CCC's desired policy (Borcherding, 1976).

46. Westbury (2001) points out that the number of pages of the U.S. Federal registry for 2000 was 83,000, somewhat less than the all-time high of 87,000 in 1980. He interprets this as indicating that the regulatory impact of government is nearly as high today as at its 1980 high point. That federal government budgets have grown a great deal since 1980 does not seem to enter his calculation.

47. The F statistic for the Wald test of whether the two coefficients equal zero is 4.48.

48. In some cases the rise has been dramatic, e.g., U.S. rising from 2.6–4.6% in 1970 to 8.9% in 1998; U.K. rising from 2% to 13%; Italy from 10.7% to 27.8%; and France from 3.9% to 14.9%.

49. Few doubt that since 1900 regulation has grown not only absolutely but also relative to the federal budget [prior to 1990 only state and local governments mattered (Hughes, 1991)]. We are reasonably confident that future research will look at the twentieth century as the regulatory century and think of Wagner's law in terms of regulation as well as in terms of government budget expenditures.

REFERENCES

Abizadeh S., and Gray J. (1985). "Wagner's Law: A Pooled Time-Series, Cross-section Comparison." *National Tax Journal* 38 (2): 209-218.

Ades, Alberto and Glaeser, Edward. (1999). "Evidence on Growth, Increasing Returns, and the Extent of the Market." *Quarterly Journal of Economics* 114 (3): 1025-1045.

Atkinson, A. B., Rainwater, L. and Smeeding, T. M. (1995). "Income Distribution in OECD Countries." *Social Policy Studies* 18, OECD.

Aubin, C., Berdot, J., Goyeau, D., and Lafay, J. (1988). "The Growth of Public Expenditure in France." In *Explaining the Growth of Government,* Lybeck, J. A. and Henrekson, M., eds., pp. 201-230. Amsterdam: Elsevier Science Publishers (North-Holland).

Barro, Robert. (1981). "Output Effects of Government Purchases." *Journal of Political Economy* 89 (6): 1086-S1121.

Barro, Robert. (1990). "Government Spending in a Simple Model of Endogenous Growth." *Journal of Political Economy* 98 (5) Part 2: S103-S125.

Barro, R. J. and Sala-I-Martin, X. (1995). *Economic Growth.* New York: McGraw-Hill.

Baumol. W. J. (1967). "The Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis." *American Economic Review* 57 (3): 415-426.

Baumol, W. H. (1993). "Health Care, Education and the Cost Disease: A Looming Crisis for Public Choice." *Public Choice* 77 (1): 17-28.

Beck, Morris. (1979). "Public Sector Growth: A Real Perspective." *Public Finance* 34 (3): 313-356.

Bergstrom, T. C. and Goodman, R. (1973). "Private Demands for Public Goods." *American Economic Review* 63 (3): 280-296.

Berry, W. D. and Lowery, D. (1984). "The Growing Cost of Government: A Test of Two Explanations." *Social Science Quarterly* 65 (3): 735-749.

Bird, R. M. (1971). "Wagner's 'Law' of Expanding State Activity." *Public Finance* 26 (1): 1-26.

Bohl, M. T. (1996). "Some International Evidence on Wagner's Law." *Public Finance* 51 (2): 185-200.

Borcherding, Thomas and Deacon, Robert. (1972). "The Demand for the Services of Non-federal Governments." *American Economic Review* 62 (5): 891-901.

Borcherding, Thomas E. (1976). "The Coastal Plan as a Statewide Zoning Ordinance." In *The California Coastal Plan: A Critique,* Bardach, Eugene et al., eds. San Francisco, CA: Institute for Contemporary Studies.

Borcherding, Thomas E. (1977). "The Sources of Growth of Public Expenditures in the U.S., 1902–1970." In *Budgets and Bureaucrats: The Sources of Government Growth,* Borcherding, T. E., ed. Durham, N.C.: Duke University Press.

Borcherding, Thomas E. (1985). "The Causes of Government Expenditure Growth: A Survey of the U.S. Evidence." *Journal of Public Economics* 28 (3): 359-382.

Bradford, D. F., Malt, R. A., and Oates, W. E. (1969). "The Rising Cost of Local Public Services: Some Evidence and Reflections." *National Tax Journal* 22 (2): 185-202.

Buchanan, J. M. and Tullock, G. (1977). "The Expanding Public Sector: Wagner Squared." *Public Choice* 31: 147-150.

Coughlin, Peter and Nitzan, Shmuel (1981). "Electoral Outcomes with Probabilistic Voting and Nash Social Welfare Maxima." *Journal of Public Economics* 15 (1): 113-121.

Coughlin, Peter, Mueller, Dennis C., and Murrell, Peter, (1993). "Electoral Politics, Interest Groups and the Size of Government." In *The Public Choice Approach to Politics,* Mueller, Dennis, C., ed., pp. 320-343. Aldershot, U.K.: Elgar Press.

DeMuth, Christopher. (1980). "The Regulatory Budget." *Regulation* 4: 39-61.

Dinopoulos, Elias and Thompson, Peter. (2000). "Endogenous Growth in a Cross Section of Countries." *Journal of International Economics* 51 (2): 335-362.

Enelow, James and Hinich, Melvin. (1989). "A General Probabilistic Spatial Theory of Elections." *Public Choice* 61 (2): 101-113.

Ferris, J, Stephen and West, Edwin G. (1996a). "Testing Theories of Real Government Size: U.S. Experience, 1959–1989." *Southern Economic Journal* 62 (3): 537-553.

Ferris, J, Stephen and West, Edwin G. (1996b). "The Cost Disease and Government Growth: Qualifications to Baumol." *Public Choice* 89 (1-2): 35-52.

Ferris, J. Stephen and West, Edwin G. (1999). "Cost Disease versus Leviathan Explanations of Rising Government Cost." *Public Choice* 98 (3-4): 307-316.

Folster, Stefan and Henrekson, Magnus. (1999). "Growth and the Public Sector: A Critique of the Critics." *European Journal of Political Economy* 15 (2): 337-358.

Folster, Stefan and Henrekson, Magnus. (2000). "Growth Effects of Government Expenditure and Taxation in Rich Countries." *The European Economic Review* (forthcoming).

Gandhi, V. P. (1971). "Wagner's Law of Public Expenditure: Do Recent Cross-section Studies Confirm It?" *Public Finance* 26 (1): 44-56.

Ganti, S. and Kolluri, B. R. (1979). "Wagner's Law of Public Expenditures: Some Efficient Results for the United States." *Public Finance* 34 (2): 225-233.

Gramlich, Edward, M. (1985). "Excessive Government Spending in the U.S.: Facts and Theories." In *Control of Local Government,* Gramlich, E. M. and Ysander, B. C., eds. Stockholm: Almqvist and Wicksell.

Goffman, I. J. and Mahar, D. J. (1971). "The Growth of Public Expenditure in Selected Developing Nations: Six Caribbean Countries." *Public Finance* 26 (1): 57-74.

Granger, C. W. J. and Newbold, P. (1974). "Spurious Regressions in Econometrics." *Journal of Econometrics* 2 (2): 111-120.

Henrekson, M. (1988). "Swedish Government Growth: A Disequilibrium Analysis." In *Explaining the Growth of Government,* Lybeck, J. A. and Henrekson, M., eds., pp. 93-132. Amsterdam: Elsevier Science Publishers (North-Holland).

Henrekson, M. and Lybeck, J. A. (1988). "Explaining the Growth of Government in Sweden: A Disequilibrium Approach," *Public Choice* 57 (3): 213-232.

Henrekson, M. (1990). *An Economic Analysis of Swedish Government Expenditure,* Ph.D. Thesis, Gothenburg University.

Henrekson, M. (1993), "Wagner's Law—A Spurious Relationship?" *Public Finance* 48 (3): 406-415.

Holsey, Cheryl, M. and Borcherding, Thomas E. (1997). "Why Does Government's Share of National Income Grow? An Assessment of the Recent Literature on the U.S." In *Perspectives on Public Choice: A Handbook,* Mueller, Dennis C., ed. New York: Cambridge University Press.

Hopkins, Thomas D. (1996). *Regulatory Costs in Profile.* Center for the Study of American Business; St. Louis: Washington University.

Hughes, Jonathan R. (1991). *The Governmental Habit Redux: Economic Controls from Colonial Times to the Present.* Princeton, NJ: Princeton University Press.

Islam, Nazrul. (1995). "Growth Empirics: A Panel Approach." *Quarterly Journal of Economics* 110 (4): 1127-1170.

Karras, Georgios. (1993). "Employment and Output Effects of Government Spending: Is Government Size Important?" *Economic Inquiry* 31 (3): 354-369.

Kau, J. B. and Rubin, P. H. (1981). "The Size of Government." *Public Choice* 37 (2): 261-274.

Kaufmann, Daniel, Kray, Aart, and Zoldo-Lobaton, Pablo. (2000). "Governance Matters: From Measurement to Action." *Finance and Development* 37 (2): 10-13.

Kormendi, Roger C. and Meguire, Philip G. (1985). "Macroeconomic Determinants of Growth: Cross Country Evidence." *Journal of Monetary Economics* 16 (2): 141-163.

Kristov, L., Lindert, P., and McClelland, R. (1992). "Pressure Groups and Redistribution." *Journal of Public Economics* 48 (2): 135-163.

Leonard, H. B. (1986). *Checks Unbalanced: The Quiet Side of Public Spending.* New York: Basic Books.

Lutters, Randall. (1998). *The Role of Economic Analysis in Regulatory Reform.* Office of Management and Budget, report to Congress on the costs and benefits of federal regulation; Washington, D.C.: Government Printing Office.

Lybeck, J. A. (1986). *The Growth of Government in Developed Countries.* Aldershot: Gower Press.

Lybeck, J. A. and Henrekson, M. (1988). *Explaining the Growth of Government.* Amsterdam: Elsevier Science Publishers.

Mankiw, N. Gregory, Romer, David, and Weil, David N. (1992). "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107 (2): 407-437.

Meltzer, A. H. (1983). "Tests of a Rational Theory of the Size of Government." *Public Choice* 41 (3): 403-418.

Meltzer, A. H. and Richard, S. F. (1981), "A Rational Theory of the Size of Government." *Journal of Political Economy* 89 (5): 914-927.

Mueller, Dennis, C. (1989). *Public Choice II.* Cambridge: Cambridge University Press.

Mueller, Dennis, C. and Murrell, Peter. (1985). "Interest Groups and the Political Economy of Government Size." In *Public Expenditure and Government Growth,* Forte, F. and Peacock, A., eds. Oxford: Basil Blackwell.

Mueller, Dennis. (1986). "Interest Groups and the Size of Governments." *Public Choice* 48 (2): 125-145.

Mueller, Dennis. (1993). "Interest Groups and the Size of Government." In *The Public Approach to Politics,* Mueller, Dennis C., ed. Aldershot U.K.: Elgar Press.

Milanovic, Branco. (2000). "The Median-voter Hypothesis: Income Inequality, and Income Redistribution." *European Journal of Political Economy* 16 (3): 367-410.

Payne, J. E. and Ewing, B. T. (1996). "International Evidence on Wagner's Hypothesis: A Cointegration Analysis." *Public Finance* 51 (2): 258-274.

Peacock, A. T. and Wiseman, J. (1979). "Approaches to the Analysis of Government Expenditure Growth." *Public Finance Quarterly* 7 (1): 3-23.

Peacock, A. T. and Scott, A. (2000). "The Curious Attraction of Wagner's Law." *Public Choice* 102 (1-2): 1-17.

Peltztman, Sam. (1980). "The Growth of Government." *Journal of Law and Economics* 23 (2): 209-287.

Perkins, George, M. (1977). "The Demand for Local Public Goods: Elasticities of Demand for Own Price, Cross Price and Income." *National Tax Journal* 30 (4): 411-422.

Pommerehne, W. W. and Schneider, Freidrich. (1982). "Unbalanced Growth between Public and Private Sectors: An Empirical Investigation." In *Public Finance and Public Employment,* Haveman, Robert, ed. Detroit: Wayne State University Press.

Posner, Richard A. (1971). "Regulation as Taxation." *The Bell Journal* 2 (1): 22-50.

Pryor, Frederic L. (2000). "Quantitative Notes on the Extent of Governmental Regulations in Various OECD Nations." *International Journal of Industrial Organization* forthcoming, (34 pages in draft).

Public Choice Society. (2000).

Ram, R. (1986). "Causality between Income and Government Expenditure: A Broad International Perspective." *Public Finance* 41 (3): 393-413.

Rodrik, Dani. (1998). "Why Do More Open Economies have Bigger Governments?" *Journal of Political Economy* 106 (5): 997-1032.

Saunders, P. and Klau, F. (1985). "The Role of the Public Sector: Causes and Consequences of the Growth of Government." *OECD Economic Studies* 4, OECD.

Schneider, Friedrich. (2000). "The Increase of the Size of the Shadow Sconomy of 18 OECD Countries: Some Preliminary Explanations." *Public Choice Meetings,* Charleston.

Spann, Robert. (1977). "The Macroeconomics of Unbalanced Growth and the Expanding Public Sector." *Journal of Public Economics* 8(3): 397-404.

Tiongson, Erwin. (1997). "Baumol's Cost Disease Reconsidered." *Challenge* 40 (2): 117-122.

Tyler, Cowen. (1996). "Why I Do not Believe in the Cost-disease: Comment on Baumol." *Journal of Cultural Economics* 20 (3): 207-214.

Wagner A. (1893). *Grundlegung der Politischen Oekonomie,* 3rd edition. Leipzig.

Wagner, R. E. and Weber, W. E. (1977). "Wagner's Law, Fiscal Institutions and the Growth of Government." *National Tax Journal* 30 (1): 59-68.

Warren, Melinda and Weidenbaum, Murray. (1999). *The Rise of Regulation Continues: An Analysis of the Budget for the Year 2000.* Washington University, St. Louis: Regulatory Center for the Study of American Business, Report 22.

Weinhold, Diana and Rauch, James E. (1999). "Openness, Specialization, and Productivity Growth in Less Developed Countries." *Canadian Journal of Economics* 32 (4): 1009-1027.

Westbury, Brian S. (2001) "The Fed (finally) Picks up the Tempo." *Wall Street Journal* (February) 1: A-22.

Zak, Paul J. and Knack, Stephen. (2001). "Trust and Growth." *Economic Journal* 111 (2): 295-321.

Chapter 5

RULES, POLITICS AND THE NORMATIVE ANALYSIS OF TAXATION

Walter Hettich
California Sate University, Fullerton
whettich@fullerton.edu

Stanley L. Winer
Carleton University
stan_winer@carleton.ca

Abstract   Over the centuries, many authors have put forward views of what quali-fies as "good" taxation and what constitutes undesirable tax policy. Con-sensus on these issues has changed over time, depending on historical circumstances and prevailing modes of economic thinking. In this chap-ter, we look at analytical views that enjoy broad acceptance in the current literature on taxation. We call these views "rules" or "norms" of analy-sis. They represent patterns of thinking that have wide currency or that have become codified in the literature. The chapter describes eight of the most important rules or norms and then critically examines their valid-ity in a framework that makes explicit allowance for collective choice. Our critique leads us to identify several shortcomings and limitations in existing patterns of thinking.

1.   INTRODUCTION

Taxation has been a much-discussed subject in the literature on economics and in writings on the role and meaning of the state. Over the centuries, many authors have put forward views of what qualifies as "good" taxation and what constitutes undesirable tax policy. Consensus on these issues has changed over time, depending on historical circumstances and prevailing modes of economic

thinking. Even in recent decades, one can find considerable variation in views among economic authors and analysts on what constitutes desirable tax policy. In this chapter, we look at analytical views that enjoy broad acceptance by a substantial group of tax policy experts. We call these views "rules" or "norms" of analysis. They represent patterns of thinking that have wide currency, or that have become codified in the literature.[1]

The chapter describes eight of the most important rules or norms and then critically examines their validity in a framework that makes explicit allowance for collective choice. In each case, we relate our discussion to basic issues or elements of public finance. Our critique leads us to identify several shortcomings and limitations in existing patterns of thinking. We group rules or norms into two categories—those that are outcome-oriented and those that are related primarily to process. In the first set, the shortcomings derive mainly from a failure to integrate collective choice into the formal framework. In the second set, we identify different types of limitations. While political decision making is explicitly acknowledged in this case, the models of collective choice relied upon are often unrealistic or incomplete, undermining suggested reforms. In addition, those using such process-oriented rules show a reluctance to extend their analysis to specific outcomes and to provide a framework that allows for a quantitative evaluation of particular economic results.

In a brief concluding section, the chapter summarizes some major implications for normative tax analysis. We argue for a more complete approach that includes an appropriate treatment of collective choice as an integral part. The discussion makes clear that development of such a framework remains a largely incomplete and challenging task. Our concern is with the structure of basic ideas. References to the literature are not intended to be comprehensive. Additional references on the issues discussed may be found in Hettich and Winer (1997, 1999).

2.   BASIC ISSUES IN PUBLIC FINANCE

In a modern economy, governments must achieve two primary goals. They must provide public goods and services demanded by the population, and they must find ways to implement changes in the distribution of income that is generated by market forces if such changes are desired by the collectivity. Theorists have pointed out a condition under which governments could finance these tasks without imposing an "excess burden," that is, welfare losses over and above the loss due to the payment of the tax. This would require that taxation to pay for public goods be levied in relation to benefits received from the consumption of these goods. Such taxes, often called benefit taxes, would act in a way that is analogous to the role played by prices in private markets. For purposes of redistribution, taxes would have to be imposed as lump sum

levies so that taxpayers could not avoid taxation by altering their behavior, while redistributive subsidies would be given out as lump sum payments.

If public goods had the same characteristics as goods sold in private markets, benefit taxation would represent a viable solution. However, many goods provided by the public sector differ in an essential manner from private goods. It is generally considered impossible to exclude those who refuse to pay voluntarily for public services, such as defense or police protection, from consuming these services. Nor is it possible to ascertain the demand for public goods by different individuals by asking them with questionnaires since potential consumers of such goods have an incentive to understate their preferences in order to minimize their own tax payments.[2]

Problems also arise if we attempt to use lump sum taxes to finance public services. Such taxes would have to be levied on individual characteristics that cannot be changed by taxpayers in order to avoid or lower their tax payments. In practice, there are few such characteristics, and payments related to them are generally perceived as inequitable or unjust, as was revealed forcefully in an experiment with head taxes conducted in the United Kingdom during the Thatcher government. Similar problems apply to lump sum subsidies; in general, recipients can find ways to adjust their economic behavior in some manner in response to being granted such payments.

Because of the problems outlined, taxes are usually assessed as compulsory levies, and individual tax payments have no direct relation to the number of units of a public good supplied to any taxpayer.[3] The separation of tax payments from decisions concerning the provision of public goods leads to excess burdens when individuals adjust their behavior to reduce their tax liability while still enjoying the benefits of public services.[4] Such excess burdens or deadweight costs also arise from redistributional policies that draw on resources raised with compulsory taxes and that provide subsidies that are not of a lump sum nature.

The need for compulsory, non-benefit taxation to finance the activities of the public sector requires allocation mechanisms that differ from those used in the private sector. Choices on what public services to provide, on how much of them to produce, and on how to pay for them must be made in a collective manner. Similarly, collective choice mechanisms are required to determine the degree of redistribution and the manner in which it will be financed. In democratic societies, allocation choices for the public sector are made through voting, and through the actions of elected representatives, or of officials to whom the political representatives delegate the power to make particular decisions concerning the use of scarce resources. Rules and norms of taxation must be evaluated in such a broader context, one that acknowledges the reasons for compulsory taxation outlined above together with the collective nature of existing political institutions that must be relied on to make fiscal decisions.

3. RULES OR NORMS IN RELATION TO THE BASIC ELEMENTS OF PUBLIC FINANCE

The discussion in the preceding section has identified three basic elements or issues of public finance: (1) separation of taxing and spending and its implications for the socially efficient use of resources; (2) determination of redistribution through the fiscal system; and (3) the necessity for non-market or collective choice mechanisms to allocate public sector burdens and benefits. The literature on rules or norms of taxation represents an attempt to codify strategies of analysis that deal with these essential elements. In this section we outline several of the most important tax rules or norms, indicating how each addresses one or more of the basic elements of public finance.

The literature on rules can be divided into two broad categories, depending on how it deals with the third element. One strand of the literature seeks solutions assuming that there is a planner who can bypass the necessity for collective choice. It is outcome-oriented, looking for detailed policy prescriptions to deal with issues arising from the other two elements, while abstracting from the necessity for collective action. We shall refer to rules arising from this approach to fiscal analysis as planning or outcome-oriented rules.

A second strand of the literature deals with collective choice allocation mechanisms as a central concern. Work in this category often focuses mainly on the nature and design of the mechanisms themselves, rather than on the detailed outcomes arising from them. If particular policies are discussed, they are seen as examples to illustrate the functioning of the process that is of importance. We shall refer to rules in this tradition as being primarily process-oriented.

3.1. Outcome-oriented Rules

3.1.1. Lump Sum Taxation as a Standard of Reference

Economists have developed a widely accepted approach to measuring the welfare losses arising from the separation between taxing and spending. As pointed out earlier, lump sum taxes differ from other levies by having no announcement effects. Since they are imposed on characteristics of taxpayers that the latter cannot change (or change only at very high costs to themselves), they do not cause a reallocation of effort or resources at the margin. The difference becomes clear if we compare a head or lump sum tax to an income tax. Economic responses to the latter are quite possible: Taxpayers can reduce their work effort and consume more leisure if they are taxed, thus reducing the utility loss from a particular tax. On the other hand, if anyone who is alive must pay the tax simply by virtue of existing, no marginal adjustments are feasible that would allow tax avoidance.

Few economists would recommend lump sum taxes as a realistic way of raising large amounts of revenues. Such taxes require that governments be indifferent to the status of taxpayers no matter how badly off they may be. Rather, this tax is used as a conceptual device to isolate the changes in allocation caused by other types of taxes. The underlying reasoning is quite straight-forward. If we can determine resource allocation with a head tax and resource allocation with some other tax, holding the amount of revenues collected constant, a comparison between the two situations must isolate any effects that are due to marginal changes in taxpayer behavior induced by the non-lump sum or distortionary tax.

Let us assume, for example, that a fixed amount of revenues is to be raised from a particular consumer, either by a head tax or by imposing a per unit tax on one of the goods purchased by the consumer. It can be shown rather readily that after-tax utility of the consumer will be higher if the given amount is raised by the head tax, since the latter does not affect relative prices and does not induce a substitution at the margin between the taxed item and non-taxed goods.[5]

Given appropriate assumptions about the preferences of taxpayers, we can determine the difference in after-tax utility levels in the two situations. It is then possible to use consumer surplus measures to assign a monetary value to this difference, although actual measurement may not be a simple matter, and may involve complex issues of estimation and calculation.[6] We call this value a measure of the excess burden or deadweight cost caused by taxation.

Not all effects associated with the provision of publicly provided goods are captured by an analysis of excess burden defined in relation to lump sum taxation. An additional problem arises with mixed goods, such as education, that have both a public and a private component. For such goods, individuals can adjust the number of units consumed, while for purely public goods, the level of consumption must be the same for everyone. If units of the mixed good are provided free of charge, consumers will demand more units than they would with efficient pricing of the good. Use of lump sum taxation as a basis of comparison will not capture welfare losses caused by inefficient pricing of mixed goods.

Welfare analysis based on lump sum taxation furthermore assumes a fixed public budget, since the comparison is made for a given amount of tax collection. If we want a standard that takes taxing as well as spending into account, we need to allow for the level of expenditures to be determined endogenously. Traditional analysis of welfare losses based on lump sum taxation does not consider efficiency losses arising from a non-optimal level of public spending.

The popularity of lump sum taxes as a conceptual device relates to the logical simplicity of the argument and to the fact that differences from the standard

of efficiency can be measured in monetary terms. It appears that excess burdens can be determined in this manner without reference to the other vexing basic issues, namely redistribution and collective choice. We shall examine at a later point whether this simplicity is in fact a justified perception, or whether it is more apparent than real.

3.1.2.   Minimization of Excess Burdens   A widely accepted rule of analysis states that a tax system is efficient if it minimizes the total excess burden of raising a given amount of revenues. This rule is an implication or extension of the use of lump sum taxation as a standard of reference.

Let us assume that the government has several well-defined tax bases at its disposal, and that it intends to assess taxes on them in such a manner that measured welfare losses are as small as possible in total. This will be achieved if tax rates are adjusted so that marginal welfare losses per (marginal) dollar of revenue raised are equal across tax bases.

A rather interesting application of this approach is the inverse elasticity rule associated with Ramsey (1927). Assume that we are dealing with a sales tax imposed on different available commodities. Minimization of excess burdens then implies that we should apply higher tax rates on commodities having a relatively inelastic demand in the relevant range of the demand function than to commodities with more elastic demands, so as to raise a given total revenue while avoiding, as far as possible, the excess burdens associated with the substitution away from commodities whose after-tax price has risen.[7]

In its simplest version, minimization of total excess burden abstracts from concerns of redistribution and collective choice. More complex versions of this rule envision a planner who uses distributional weights derived from a welfare function given from outside the conceptual framework.[8] In such a context, the planner attempts to maximize social welfare. To achieve this, he or she will equalize distributionally weighted marginal excess burdens per dollar raised across available tax bases. This more general approach, now called the theory of Optimal Taxation (OT), has become established through the work of Ramsey, Mirrlees (1971) and others as the most influential normative approach in taxation.

While the social planner or OT approach allows incorporation of a second basic issue (redistribution), it does so at the expense of practicality. In actual policy contexts, well-defined welfare functions are not available, and it may be difficult to determine even in an approximate fashion what the prevailing consensus is regarding distributional weights. One should also note that the planner model completely skirts the third basic issue—the necessity for collective choice—since it describes a standard of reference drawn up without regard to the costs of collective choice. We shall return to this point later in the chapter.

3.1.3. Tax Neutrality

There is another reason, besides the difficulty of identifying distributional weights, that substantially reduces the practicality of the optimal tax approach. Those concerned with the reform of particular tax systems point to the heavy informational requirements of OT. Since optimal tax plans take full account of the general equilibrium structure of the economy, they tend to be highly complicated and complex. To develop a comprehensive OT blueprint of the tax system, the social planner needs knowledge of preferences, endowments and technology for all participants and sectors in the economy, as well as knowledge of distributional weights. How this information is to be acquired by elected politicians who are in charge of policy making is not addressed.

If we restrict the analysis to commodity taxation, the primary need is for information on demand functions and commodity characteristics. Stern (1987) has discussed the conceptual problems involved in the generation of such data:

> The derivation of the appropriate set of commodity taxes requires information concerning patterns of complements and substitutes that is very difficult to extract from the data. Our attempts to extract it will require specifications of functional forms, which … may have a profound effect on the recommendations. As Deaton … observes: "In consequence, it is likely that empirically calculated tax rates, based on econometric estimates of parameters, will be determined in structure, not by the measurements actually made, but by arbitrary, untested (and even unconscious) hypotheses chosen by the econometrician for practical convenience" (1987, p. 51).

The problem is particularly acute in developing countries, where the necessary information systems are largely absent, but it also exists in more developed nations, where planners face a bewildering array of different goods and constantly changing market conditions.

Suggestions in the tax literature for dealing with the information problem center on rules of thumb, or simplified guidelines such as tax neutrality, Neutrality here means that all taxable activities should be treated equally by the tax system (that is, taxed at the same effective marginal rate) in order to avoid a far as possible the excess burdens that will arise as taxpayers substitute towards relatively lightly taxed activities. As one writer (Gillis, 1989) has put it:

> While not nearly as intellectually satisfying a guide to tax policy as "optimal taxation," neutral taxation is to be preferred as a benchmark until such time as analysts are able to identify optimal departures from neutrality in real world policy settings, and until such time as administrative capacities are equal to the task of operating necessarily complicated optimal tax structures (1989, p. 515).[9]

In other words, those who advocate tax neutrality recognize that it is less efficient than a properly specified optimal tax blueprint, but argue that a neutral system will be more efficient than any feasible OT system that is badly implemented. We return later to whether or not this conclusion is justified.

3.1.4. Harmonization: International and Interregional Neutrality

Neutrality in an international or interregional context is often referred to as harmonization or fiscal coordination.[10] The application of the approach is similar in both the international and the interregional contexts, though specific policy recommendations consistent with neutrality within a federal state are generally more detailed than those considered to be feasible in the international context. For convenience, we shall confine the discussion to the international case.[11]

The problem in the international context is to devise simple rules that allow for the financing of public goods as well as for redistribution within and between countries, while taking account of the possibility that firms and consumers will move across international borders to minimize their tax liabilities. When tax payers are mobile, there is a danger that international competition will lead to the bidding down of national fiscal systems to a level at which the only taxes that can be collected are those that support the business-oriented services which multinational firms are prepared to pay for.

In this sort of situation, moreover, the presence of fiscal externalities in domestic public decision making may lead to suboptimal levels of public services even if some non-benefit elements in the tax system remain. Externalities arises because, in the face of tax base mobility, the cost of raising revenue to pay for public services includes not just the full domestic cost (including excess burden), but also the loss of revenue due to tax base mobility. A national government will take this extra cost of domestic programs into account when setting national tax rates, but will not consider the offsetting gain in taxable activity that may accrue to other countries. From a global perspective, the overall result is an equilibrium with an inefficient level of public services, where the public sector in each country is usually too small.

Ingenious rules have been formulated to permit the financing of public services while preserving intra- and inter-nation equity, all in the face of internationally mobile factors, commodities and services. To the extent that these rules succeed, they do so because they reduce the usefulness of international mobility as a way of avoiding tax liabilities.

One system of interjurisdictional harmonization that has developed over many years (see Musgrave, 1991, who in part credits Seligman, 1921) is one that combines the residence principle for the taxation of international income from capital for residents of a country with source taxation of nonresidents and a foreign tax credit by the nonresidents' home country.[12] Under the residence principle, the capital income of citizens is taxed in the same way no matter where in the world it is earned. This is an example of tax neutrality in the international setting. Since tax payments do not depend on where they are earned, intra-nation equity is preserved between those who have and those who do not have foreign source income. For the same reason, the capital owned

by residents will be efficiently allocated around the world (from a national perspective), a situation often referred to as capital export neutrality.

The taxation at source of income earned domestically by nonresidents preserves inter-nation equity, by giving the country in which resource and other rents are earned the first 'nibble'. And, at the same time, the foreign tax credit extended to nonresidents by their own government eliminates double taxation of their foreign source income, in effect preserving the residence principle despite taxation at source of nonresidents.[13]

For commodity taxation, the analogue to the residence principle is the destination principle, under which purchases by residents are taxed the same no matter where they are made (i.e., imports and domestic purchases are taxed alike), while purchases by nonresidents are not taxed (i.e., exports are not taxed). The application of the destination principle insures that prices received by foreign and domestic producers selling into the same market are identical, thus preserving production efficiency even though different countries may have different tax rates.

It is important to note that in this residence-destination tax regime, the pressure for international tax competition is much reduced. Each country's residents pay the same tax no matter where they transact, while nonresidents are (under a foreign tax credit system) largely unaffected by the domestic tax system. If one country raises its tax rates unilaterally, this may lead to domestic political protest, but it will not lead to harmful international capital or commodity tax arbitrage.[14] Moreover, because a harmonized international tax system attenuates the incentives for mobile tax bases to move away from relatively higher tax jurisdictions (since they still pay the same, domestically determined tax), it also reduces fiscal externalities.

Like tax neutrality, the rules for international harmonization that have been worked out over the years are cognizant of the inefficiencies that may arise because of the separation of spending and taxing and the need for redistribution. The role of collective choice in this literature will be addressed later.

3.2.   Process-oriented Rules

3.2.1.   The Comprehensive Tax Base and Horizontal Equity   There is a large body of literature that relates to horizontal equity and the comprehensive income tax base. The major proponent of this approach was Henry Simons who published his work in the 1930's and 1940's. In his classic book *Personal Income Taxation* (1938), he spelled out the major arguments for levying taxation primarily on income, and for using a comprehensive definition of income in determining the tax base.

While Simons' work on income taxation extended a tradition of analysis originally developed by Haig and Schanz, he added an important new element. He advocated a comprehensive tax base as a way of limiting government

interference in the economy. He believed that the tax base could be defined according to logical principles that would command broad support once they had been aired widely in public discussion. His definition of the appropriate base—the change in net wealth plus the value of consumption during an accounting period—was thus offered as a process-oriented rule that would circumscribe government intervention in the private economy.

The definition of the comprehensive tax base depends on the related concept of horizontal equity. Simons argued that those with equal ability to pay taxes should be assessed equal tax payments. He saw this principle as a way of implementing justice in taxation and believed that it would have wide support among taxpayers, leading to a tax system that would be perceived as fair among the population.

The concepts of horizontal equity and of the comprehensive tax base made a lasting impression on economic writings related to taxation and also had a considerable impact on the legal profession. Simons had rejected a utility-based analysis and directed the focus to implementation of the comprehensive tax base in a manner having a direct counterpart in accounting practices. Many writers followed this lead. The result was a voluminous literature dealing with problems of implementation and an extensive debate on what should and could reasonably be included in a comprehensively defined base.

The work by Simons and by many later writers, such as George Break, Joseph Pechman and Richard Musgrave, had an important influence on attempts to reform existing tax laws in the United States, Canada and several European countries. The apogee of this type of analysis was probably reached in the Report of the Canadian Royal Commission on Taxation (1966), that spelled out in volume after volume just how comprehensive income taxation should be implemented at the federal level in Canada. In the United States, ideas in this tradition repeatedly influenced policy discussion, with the most recent example being the debate leading up to the 1986 Tax Reform Act.[15]

While ideas from this tradition continue to influence policy, they have lost favor in recent years among theoretical economists with the development of the optimal tax approach. The Simons tradition also conflicts with more recent writings on collective choice that treat the tax system as an equilibrium outcome of the political process. We shall return to these differences in our critical examination of existing norms.

3.2.2.   Limiting the Power to Tax   A second process-oriented approach also focuses on the definition of available tax bases, but it reaches conclusions that contrast starkly with those arrived at in the Simons tradition. The reason for this relates to differences in assumptions concerning the motives and actions of government. The literature on limiting the power to tax starts from the premise that public decision-makers attempt to maximize total

revenues that can be extracted from the private sector. A comprehensively de-
fined tax base would provide increased opportunities to those who have such
Leviathan-like motives. Unlike Henry Simons, who believed that the political
process would set appropriate targets for budget size after public discussion,
writers using this approach argue that determination of budget size must be the
primary focus of the analysis.

Geoffrey Brennan and James Buchanan (1980) provided the classic state-
ment of the reasons for limiting the government's power to tax. They argue
for a tax constitution that would restrict fiscal decision-makers to narrowly de-
fined bases. Economic activities that are relatively elastic with regard to tax
rates are preferred, since their inclusion in the constitutionally determined tax
base allows taxpayers to adjust behavior, and thus to reduce the size of the total
budget by escaping into non-taxed alternatives. While such avoidance would
create welfare losses, Brennan and Buchanan believe that a tax constitution of
this nature would lead to greater overall efficiency, since it limits the excessive
growth of public budgets and of the public sector.

Like the Simons' approach, the literature on limiting the power to tax puts
the third basic element in tax analysis, namely the need for collective action, at
the center of the argument. Detailed examination of the other two elements is
subordinated to a review of the implications of political choice, though the ul-
timate concern of the approach is with the total deadweight costs of the public
sector. However, as with the argument for a comprehensive tax base, the ap-
proach depends crucially on the model of collective behavior that is chosen. As
will be shown in a later section, the use of alternative and, we will argue, more
realistic models leads to quite different conclusions, even if political choice
remains the centerpiece of the analysis.

3.2.3.   The Generality Principle   In a recent contribution that has not
yet been widely discussed by other scholars, Buchanan and Congleton (1998,
Chap. 8) have returned to the concern of Simons with the problems of majority
rule. Their analysis of the operation of majority rule is more explicit than in Si-
mons' work, however, and it includes an examination of the dynamic character
of democratic politics.

Buchanan and Congleton argue that rent seeking and static welfare losses
are likely to occur since majorities often do not internalize the losses suffered
by minorities. Over time the losses due to majoritarian exploitation of minori-
ties may become steadily more serious. Because of the inherent tendency of
majority politics to foster vote cycling and instability in the struggle over dis-
tributive shares, the long-run rate of growth may decline, unless appropriate
constraints are placed on political outcomes. This will be so because any cur-
rently successful coalition may find it desirable to raise tax rates even higher
than merely static political optimization would suggest, because it realizes that

it will probably not be the majority tomorrow, and that it cannot fully capture the future gains from growth enhancing policies.

Interestingly, the solution advocated is similar to Simons' broad base income tax, the purpose of which was to prevent government from dipping into great incomes with a sieve, as Simons (1936) put it. In this more recent contribution to the process-oriented literature however, the suggested solution involves the application of a Generality Principle, or principle of non-discrimination, under which all citizens are to enjoy equal quantities of public services, and pay taxes according to a flat or uniform tax system on a broad base that does not permit economic activity of particular groups to be singled out for 'special' treatment. The intention here is to avoid distributional conflict, and the vote-cycling and associated economic waste that results. In order to do so, Buchanan and Congleton go so far as to argue that it is necessary to eschew the use of exemptions that remove low income taxpayers from the tax rolls, since this invites political conflict over who is to be exempted.

3.2.4.   Time Consistency: Neutrality over Time   The last process-oriented approach we shall consider suggests the use of independent agencies (much like independent central banks), policy rules, or social contracts to distance the setting of policy from the day-to-day vagaries of democratic politics. In this literature, the focus of concern is on the inefficiency that may arise in a dynamic context when contemporary governments, perhaps acting on behalf of majority coalitions, engage in discretionary fiscal and other polices that are not consistent over time.[16]

A policy is not time consistent if it requires a course of action today that will subsequently become undesirable. It is argued that the inability of governments to commit to consistent policy over time will result in a loss of social welfare compared to a situation where governments are prevented from adopting (discretionary) policies based on period by period political optimization. In other words, the concern here is with the welfare losses that may occur in the absence of policy neutrality over time.

An example of how the inability of majority coalitions or governments to commit to a time consistent policy reduces welfare may be useful (see Kotlikoff et al., 1988 and Fischer, 1980). Suppose an ex ante efficient policy (in the absence of lump sum taxation) in an intergenerational context involves a low rate of taxation now and in the future to encourage saving when people are young. Such a policy will not be time consistent without some special institutional arrangement, however, that commits the older generation to refrain from taxing capital once it has been invested. In the absence of a credible commitment not to tax capital when they are, as a group, old, the young will distort their current saving patterns in anticipation of future taxes. The result will be

an economy with sub-optimal saving, lower real growth, and high capital taxation.

The usual policy recommendation emerging from consideration of the time inconsistency problem in the setting of monetary policy is a call for an independent central bank, insulated from day-to-day political pressures, with the power to determine the rate of monetary growth. Although consistency issues regarding taxation are similar to those for monetary policy, it is much less common to hear requests for quasi-independent tax commissions with the power to determine the structure of taxation. One should note, however, that Simons realized the connection between the two issues. He called both for the introduction of a broadly based income tax (1938) and for the adoption of a strict monetary rule (1936) in order to avoid the uncertainty and inefficiency created by what he considered the whims of the political process.

The analysis by Buchanan and Congleton is tailored to deal with much the same problem as the time inconsistency literature. One may thus reasonably ask, whether the flat tax without exemptions advocated by them would serve as an analog to an independent central bank. (Given Simons' work, the same question could also be asked for the comprehensive tax base.) Unfortunately, ready answers are not to be found in the literature, which fails to consider the strong existing parallels between the two situations.

4.   EVALUATION OF RULES OR NORMS

In the discussion of various norms, we have repeatedly referred back to the three basic elements of public sector analysis. We shall argue that the shortcomings of various rules relate to the partial nature of these rules when they are evaluated in relation to a comprehensive analysis of all three elements. Although particular norms may serve as acceptable analytical tools within a more narrowly defined framework, they are revealed as incomplete in a more encompassing analysis.

Lump Sum Taxation: The first norm—lump sum taxation—abstracts from the need for budget determination through collective choice. Budget size is taken as given. A more complete analysis would compare observed taxation to an ideal standard that allows for the determination of the overall budget in political equilibrium, as well as for the most efficient way of raising tax revenues.

How could we envisage such a more inclusive standard? One approach that has been suggested is to extend the theory of optimal taxation to include the determination of budget size, following Atkinson and Stiglitz (1974). This requires reformulation of the Samuelson conditions for the efficient provision of public goods, so as to allow for the use of distortionary or non-lump sum taxation. Such an approach will not solve the problem, however, if the need for

collective choice is admitted as a starting point. In that case, the standard of reference will have to emerge from an efficiently functioning collective choice process. One may note that it is unlikely that such a process would include the use of lump sum taxes, since they do not appear to represent an acceptable policy tool in modern democracies.

There are some theoretical difficulties in determining the exact nature of the tax system that would emerge from an efficiently operating political system. If we use Wicksell's (1896) work as a starting point, we may propose a system reached with unanimous decision-making (or with some approximation to unanimity) as the standard. Or we may turn to Lindahl's (1919) writings and the subsequent work on the nature of Lindahl equilibrium (see for example Foley, 1977) for a more formal development of this approach. Alternatively, analysis can start from the literature on probabilistic voting where politically competitive equilibria have been described in some detail. Such equilibria do have existence and are stable. In addition, they can be shown to be Pareto-optimal in some circumstances (Coughlin and Nitzan, 1981), suggesting that they have the necessary characteristics to serve as a standard of reference for tax analysis.[17]

Because of the difficulties of implementing a more inclusive standard of comparison, we may decide to continue to rely on lump sum taxation in determining welfare losses. If we do so, it will be imperative to develop a better understanding of the biases that use of lump sum taxation introduces into the analysis of tax efficiency.

Minimization of Excess Burden: A tax reform imposing policies based on equalizing marginal excess burdens (per marginal dollar collected) across tax bases will lead to Pareto improvements only if we postulate a framework where collective choice has no explicit role. This can best be seen if we consider the question using an alternative standard, namely an equilibrium reached in a world with competitive parties and free political entry, where political decision-makers continually adjust policies so as to maximize expected votes in the next election. Policy equilibria in such a setting may be Pareto-optimal, as noted earlier.

However, decision-makers in such a competitive political system will not equalize the *unweighted* marginal welfare losses across tax instruments; they will also have to consider the impact of a marginal dollar raised on the probability of getting voter support. Since different taxpayers, or taxpayer groups, will react differently to a given tax payment and welfare loss, and since various tax bases are associated with differing incidence patterns, it would not be politically rational to equate unweighted marginal welfare losses in equilibrium. Studies that have found higher such losses for capital taxes than for wage-based taxes may thus have uncovered a pattern that could be both ratio-

nal and economically efficient if viewed in a more inclusive, political economy setting.

This conclusion contrasts with the arguments of several authors, especially those who favor tax neutrality as a guide to policy, who see unequal marginal welfare losses across different bases as a source of grave inefficiency. (To complicate matters further, it must be pointed out that the existence of differences in marginal welfare losses does not indicate by itself that an efficient political equilibrium has been achieved. Such differences could possibly also reflect the influence of an inefficient policy that arises from non-competitive elements in the political process. Pareto optimality will only be reached if the political process is truly competitive.)

One may perhaps reply that it is difficult, or even impossible, to determine the optimal political weights that would be used by a government in a world with strong competition among parties and probabilistic voting. While this may turn out to be the case, it should be noted that little work has been carried out so far in order to understand or estimate weights of this nature.[18] Given such weights, it would then be possible to determine if there are Pareto gains still to be realized.[19] Furthermore, probabilistic voting represents only one formal approach to this question. The theoretical difficulties just outlined arise from the existence of a costly collective choice process, not from the use of a particular voting model. While unweighted marginal welfare losses (or losses adjusted for distributional objectives reflected in a welfare function) may provide a proper guide to an efficient tax system in a world where decisions are made by a benevolent planner, they cannot play the same role in a context that allows for the necessity of a collective choice mechanism and less than perfect equality of effective political influence.

Neutrality: Neutrality in the static context is intended as a reasonable guide to policy when optimal tax plans are too complicated to design and administer. Since neutrality rules have emerged from a framework that is not cognizant of the need for collective choice, one may reasonably ask if neutrality is still a useful guide to policy in a more complete, political economy framework.

It is easy to see that the information problem for policy makers becomes worse in any political economy setting. The political strategist must have knowledge of all relevant political margins governing voting behavior, in addition to the information about economic margins required by a traditional OT planner. A full solution to the problem of optimizing political support requires knowledge of how changes in the welfare of different voters affect the probability of voting, as well as how taxation affects economic behavior. Only then can the tax system be adjusted correctly to favor those voters who are more likely to offer the party support at the polls. It therefore appears that the argument for neutrality is stronger when the existence of collective choice is explicitly acknowledged.

However, the historical debate points in a rather different direction from neutrality as a solution to the information problem, for both economic policy makers and party strategists. The feasibility of social planning in the face of large information requirements is a classic question in the history of economics. The traditional debate was concerned primarily with the choice between centralized planning and the use of markets. Among the most influential ideas in the debate were those advanced by Hayek (1945), who argued strongly that only decentralized markets could solve the immense task of processing the information necessary to reach efficient economic outcomes. This approach is in contrast to that taken by advocates of tax neutrality, who have retreated from optimal taxation in order to deal with the information problem, while still preserving a command and control approach to policy making.

The historical debate points in a rather different direction from the use of neutrality as a means of economizing on information costs.[20] It suggests that a more effective approach may be to decentralize policy making into separate, semi-independent areas, while mobilizing special interest groups to provide valuable information as part of their attempts to influence policy outcomes. One may note that the most commonly used OT formulation also subsumes a segmentation or decentralization of policy by separating taxation from expenditures, although authors do not generally justify this assumption by making reference to the information question.

The study of policy making in modern societies indicates that decentralization of policy areas is a common feature of democratic government. In the United States, Canada and Europe, for example, decisions on taxation and expenditures are taken separately at the political level, and implemented by different administrative bodies, while special procedures, such as annual budget resolutions or cabinet directives, are used to maintain broad overall coordination. As far as taxation is concerned, further segmentation of policy making and administrative organization tends to occur in accordance with particular fiscal instruments or major tax bases. Moreover, instructions to tax commissions and tax reform policies are usually directed at selected parts of the revenue structure.

While the apparent lack of coordination that may result is often decried by economic analysts, this lack may in fact represent a rational response to information problems associated with complex policy choices. To fully understand the strengths and weaknesses of existing, decentralized policy processes, which is a prerequisite to the conclusion that neutrality is the best that can be hoped for, it will be necessary to define and examine the benefits and costs associated with existing methods of decentralization and policy segmentation, and to relate them to the provision and processing of economic and political information necessary for electorally effective policy.[21]

Harmonization: The implicit assumption lying behind a harmonization regime is that uncoordinated government actions reduce economic welfare even when each government strives to maximize the welfare of its own citizens. One can say that the harmonization literature is more complete than the other outcome-oriented approaches discussed above since it does, implicitly, contain a theory of government, albeit a rudimentary one. However, as Mueller (1998, p. 180) and Breton (1996) ask, why should we expect competition between governments always to produce inefficiency while competition between private sector agents produces efficiency? And if intergovernmental competition does reduce the possibility for inefficient government behavior, then harmonization, which reduces the extent to which states must compete for taxable activity, will come at a cost in terms of government performance. The assumption that every government behaves benevolently, and the absence of a theory of interjurisdictional competition, are the Achilles heals of the harmonization literature.

It should be noted that no race to the bottom, much feared by writers on international harmonization, will ensue if governments supply goods that their citizens want, and if taxes are not unduly coercive. If the mobility of tax bases is a serious problem, adjusting the mix of taxes so as to rely more heavily on inelastic bases such as property may be useful. Residents can avoid such taxes by moving, but only if they also give up the benefits of public services these taxes finance (Mueller, 1998). In other words, not only are the harmonization rules based on a conceptual framework that contains at best a rudimentary theory of government, but they may also be based on an unduly pessimistic assessment of the ability of each nation-state to overcome the problems of raising revenue in the face of the declining costs of international transactions.

In their current state, neither the traditional economic literature nor the literature that is cognizant of collective choice appears to offer practical guidance in dealing with the trade-off between the gains from coordination and the losses from attenuating intergovernmental competition as a disciplining device.[22]

Horizontal Equity, the Comprehensive Tax Base, and the Generality Principle: Although Henry Simons viewed the comprehensive tax base as a way of limiting government discretion in determining and changing tax rates, he did not place the discussion within a formally developed framework of collective choice. As a result, it is not clear why political actors would adopt a truly comprehensive base or why they would choose horizontal equity as their main policy criterion. The same question can be posed concerning the proposal of Buchanan and Congleton for a nondiscriminatory flat tax without exemption.

The voluminous literature on the comprehensive tax base has focussed almost exclusively on problems of implementation and on theoretical arguments

about whether income or consumption would provide a better basis for taxation. While reformers starting from this tradition have influenced public discussion and public policy to a considerable extent, they have not succeeded in having their agenda accepted fully. The reasons for this lack of success can be understood more readily if we consider the choice of tax policy in the context of political equilibrium.

Let us imagine a political system where both the party in power and the one in opposition propose policy platforms so as to maximize expected voter support, while being uncertain of how voters will react to particular proposals. Voters, in turn, will try to maximize net benefits from the public sector, putting a positive value on public goods and services and reacting negatively to the payment of taxes and to the welfare losses arising from taxation. In such a system, political decision-makers will achieve their objectives if they design an equilibrium tax structure that equalizes the change in opposition per marginal tax dollar raised across all taxpayers or taxpayer groups.

In formulating their platforms, parties face a difficult balancing act. On the one hand, they want to create a tax system with as much differentiation in the treatment of taxpayers as possible in order to minimize total opposition. On the other hand, they face information, administration and monitoring costs that increase as more differentiation is introduced. Such costs reduce the ability to provide public services and, for this reason, lead to a loss in expected support. The equilibrium tax system must represent a compromise between these opposing forces. Differentiation is reduced by grouping taxpayers into rate brackets and by combining disparate activities into large bases. However, some of the lost ground is regained by using special provisions, such as exemptions, deductions and exclusions that provide some measure of differentiated treatment with regard to effective tax rates, even with the existence of large omnibus bases.

An analysis of this nature suggests that democratic governments operating in a competitive political environment will not voluntarily implement a tax program corresponding to the one advocated by Henry Simons. The reason is twofold. First, a broad base income tax without special provisions may make people worse off compared to the outcome that can be obtained in a competitive political system, thereby reducing the support that political parties can expect at the polls. Second, such a system makes no allowance for responses to differences in effective political influence among individuals and groups in society, differences that are tolerated or sometimes even encouraged. While horizontal equity may enter into the government's calculus, among other issues, to the extent that it represents a widely shared value among taxpayers, it will not be the overriding criterion in the fashioning of a tax structure that is consistent with political equilibrium.

The objection to the Generality Principle is essentially the same as that to the broadly based income tax. In the end, the issues are empirical as well as theoretical—how does the political system operate; can it be altered so as to intensify the economic benefits from political competition, or must it be constrained to avoid the worst features of interest group politics?

Limiting the Power to Tax: The literature on limiting the power to tax places the government's motives and actions at the heart of the analysis. Authors writing in this tradition propose a specific model of government behavior. They assume that those in power act as monopolists, attempting to extract as much of the real resources of the private economy as possible. To counter such designs by the taxing authority, these authors suggest a written constitution that limits those in power to the taxing of narrowly pre-defined bases.

Three specific criticisms of the assumed model of government behavior are relevant. The first concerns the nature of contracts. All political systems function within a complex system of contracts and agreements. Constitutions represent merely one among many available contract types. They differ from other forms mainly by making it more difficult to effect change, although there usually are mechanisms in place to amend their provisions.

If we look at existing political frameworks as sets of functioning contracts that can be changed at differing costs, it is no longer clear why a written constitution would make a fundamental difference, or why society would necessarily want to choose the resulting loss in flexibility with regard to the making of tax policy, since constitutional restrictions would no doubt create welfare losses and other costs of their own.

The second criticism is related to the process of adopting constitutions. Creation of constitutional restrictions has to occur within the existing set of political contracts. This is illustrated by recent far-reaching constitutional change in Canada in the 1980's and the European Community in the 1990's. A complete theoretical analysis would thus have to show how a tax constitution could arise as an equilibrium outcome of a functioning political system.

It should also be noted that the literature on the Leviathan model of government does not present an analysis of how its postulated governing structure could arise. Nor is it clear what forces would allow such a structure to remain as a stable outcome, once it had been established. Thus, the model does not satisfy the requirements for a complete discussion of collective choice, despite its emphasis on political motives and decision-making. For example, it is not apparent what determines entry into the governing group or elite, and what may limit the political power to enforce Leviathan-like policies. As a result of these criticisms, the conclusions concerning tax structure that emanate from this tradition cannot readily be applied to the study of tax systems in democratic societies.

Time Consistency: The literature is of interest because it raises again the question of how contracts of different types are to be enforced over time in a democratic society. It is not surprising that the complexity of the issue leaves many facets of the problem still to be explored.[23]

In the first place, the time consistency problem is partly an empirical issue (Sheffrin, 1989; Taylor, 1983). Clearly not all contracts are broken by governments. For example, patents are not usually abrogated unilaterally. Knowledge of the extent to which inconsistency problems actually exist is important, as noted earlier, because giving up discretion through the use of rules or independent agencies must have a cost.

One reason for thinking that time inconsistency may not be as serious a problem as the models suggest is that people in democratic societies are not powerless in opposing unwanted government actions. The legal system in most developed nations contains features that make it difficult for governments to unilaterally expropriate private property. This is even true in countries such as the United Kingdom, where there is no written constitution. Relevant property rights are not inviolable, but neither are they absent. Mobility and the organization of political opposition are other well used methods available to taxpayers that make them more difficult targets than the time inconsistency argument suggests. Indeed, it is not farfetched to say that the type of expropriation of taxpayers envisaged in models of time inconsistency amounts to the staging of a coup by the government, which is unlikely to happen in most democracies for a variety of reasons.

In any event, there are many facets of existing arrangements in democratic societies that we do not fully understand, making policy advocacy in the area a dangerous enterprise. Why for example do we observe the existence of quasi-independent central banks with authority for monetary stability, as is in accordance with the time inconsistency approach, but we do not observe in any democratic state the corresponding institution of central taxing, even though the time consistency problem with respect to the rate of inflation and the rate of capital income taxation are essentially similar?

A possible answer to the last question is that tax policy is hard to design and implement in the face of constantly changing events, a point emphasized in the discussion of the rule of static neutrality. Indeed, it is this difficulty that lies behind the advocacy of a neutrality rule. It is not hard to see why it is easier to write a contract with a central bank to carry out a program of monetary stability than it is to instruct a central tax authority on the appropriate definition of taxable income in a constantly changing economy.

If central taxing is rejected as a solution, it is tempting to look at the advice of Simons or Buchanan and Congleton regarding the definition of income and the structure of taxation to deal with government opportunism. However, in view of the ability of taxpayers to protect themselves to some extent with

mobility and legal and political action, it is necessary to consider the trade-off between the benefits from imposing such tax structures and the costs of policy inflexibility. The nature of this trade-off remains to be studied.

A further complication in the evaluation of solutions to the time inconsistency problem concerns the degree of political stability possible under rules rather than discretion. Boylan and McKelvey (1995) show that if two candidates can commit to a multiperiod consumption path and voters are heterogeneous, no majority rule equilibrium exists. The reason is that by bundling periods together, commitment creates a voting game of high dimensionality. Such commitment in their view implies randomness in economic outcomes. On the other hand, period by period discretion leads to the median voter outcome in each period given their framework, since in this case, voting is confined to one issue at a time, implying a type of structure-induced equilibrium. The argument complicates the evaluation of rules versus discretion by interjecting the question of what sort of consistency over time is compatible with an acceptable collective choice process.

5.   TAXATION, WELFARE ECONOMICS AND POLITICAL MARKET FAILURE

The evaluation of current norms of analysis suggests a need for greater comprehensiveness and integration in the study of normative taxation. While outcome-oriented rules fail to account in a satisfactory manner for the necessity of collective choice, process-oriented rules fall short because they are based on misleading or incomplete models of collective decision making. Moreover, they provide little guidance on how to empirically evaluate specific equilibrium tax policy outcomes.

Although the literature based on the planner model avoids dealing with essential questions arising from collective choice, it makes a valuable contribution by formulating a well-developed agenda for the study of policy outcomes. A more comprehensive approach must include many components of this agenda, while at the same time integrating the formal modeling of collective processes. Evaluation of outcomes remains an important task for economists, even if collective decision-making becomes the center of attention. It is essential to understand equilibrium outcomes that are produced by well-functioning political processes, and to examine how such outcomes change when imperfections become part of collective choice.

5.1.   Three Steps to a More Comprehensive Analysis

One way of constructing a more comprehensive framework is to seek guidance from the classic literature on welfare economics that was formulated for the evaluation of private markets. This well-developed body of work contains

three essential steps. First it provides an elegant analysis of why and how markets achieve a Pareto-optimal allocation of resources, an analysis that is summarized by the First Theorem of welfare economics. Examination of allocative efficiency in competitive equilibrium is then complemented by the study of market failure and of the causes leading to such failure. Finally, the literature provides a framework to measure the consequences of market failure with the help of the concept of economic surplus. Thus a dollar value can be assigned to welfare losses that arise from the existence of market imperfections.

We would argue that a more comprehensive normative analysis of taxation must include all the elements embodied in the three steps taken in welfare economics. This implies that we need a model of collective choice as our starting point that allows us to study and demonstrate the existence and stability of political equilibria and to examine the nature of specific equilibrium policies or outcomes. Probabilistic voting provides one approach that enables us to accomplish this, since it can be demonstrated that the resulting Nash equilibria between or among parties are Pareto-optimal. (For a detailed exploration of this point in the context of tax analysis, see Hettich and Winer, 1999, Chapter 4.) One should note, however, that the need for taking this basic analytical step is not tied to the use of a particular framework; rather, it arises from the fundamental nature of normative analysis itself.

Imperfections in private markets have their counterparts in failures of the political process. To take the second basic step, we must focus on the operation of the collective decision mechanism in order to identify features that cause it to operate imperfectly. The challenge is considerable. Not only must we begin by modeling a political process that leads to an optimal allocation of resources, but it is also necessary to determine specific tax policies that will be part of the political equilibrium. Once this has been accomplished, we can then extend the examination to particular imperfections in collective decision making and trace out their implications for the nature and structure of tax policies.

Few authors writing on taxation have concerned themselves with this part of a more comprehensive research agenda. The need for such work is evident however. Unless it is carried out, economists cannot present an analysis of tax policy failure that has the same force as does the well-known work on private market imperfections.

It should be noted that the optimal tax literature is also concerned with efficient taxation. However, writers in this tradition define optimality with reference to a planner who maximizes an exogenously given welfare function. Since no collective choice mechanism is incorporated into the analysis, we cannot examine equilibria that result from the operation of a political process, nor can we examine the effects of process failure within the specified theoretical framework. Optimal solutions determined in this approach would have to be imposed on a collectivity from the outside. Since such solutions are not an

endogenous outcome of the society's political arrangements, it remains unclear whether they would be consistent with the workings of collective institutions and whether they could ever represent stable policy equilibria.

The final step relates to measurement. There is an extensive literature in economics dealing with the quantitative evaluation of welfare losses created by taxation. As pointed out in an earlier section, most of this work uses lump sum taxation as a standard of comparison. The theoretical difficulties that arise concerning this approach, when it is evaluated in the context of a framework that contains collective choice as an essential element, have already been explained. The challenge for research in this area is to define and measure welfare losses in relation to a standard of reference that is consistent with the operation of the political process, or failing that, at least to learn more about the biases that the use of lump sum taxation introduces into the calculation of such losses.

5.2.   Other Approaches to the Redefinition of Normative Analysis

A key characteristic of the approach to policy analysis we have outlined above is that the status quo is always compared to allocations that can be supported as equilibria of a competitive political system. In a recent contribution, Besley and Coate (2002) also advocate this approach to policy analysis. However, the details of their argument differ from the approach suggested above in an important way.

To argue the case for a policy analysis that is confined to comparisons of political equilibria, Besley and Coate ask: will the availability of a new policy instrument lead to a political equilibrium that is Pareto-superior to the status quo equilibrium (without the new instrument)? One may argue that the appearance of a new instrument may be used in principle to make someone better off. Moreover, parties will have an incentive to adopt new instruments since the chances of electoral success are generally enhanced by increasing the welfare of some voters. Besley and Coate show, however, that the appearance of a new instrument may in fact lead to a new equilibrium in which some people are much worse off than before, so that the new equilibrium does not Pareto-dominate the status quo. Moreover, the new equilibrium may also be inferior when judged in terms of a utilitarian social welfare function.

Imagine for example that a new type of public good appears that is favored by the rich, but disliked by the poor. It is possible, depending on voting behavior, that the expected equilibrium outcome could favor the rich, and that the poor would be worse off in the new equilibrium. Preferences, voting behavior, and the nature of political competition all interact in various ways to make the outcome created by the appearance of the new instrument unclear, provided

that attention is confined to political equilibria, and that hypothetical alloca-
tions that do not represent equilibria are set aside. Another general lesson of
the Besley/Coate paper is that it matters for welfare analysis how public policy
is actually made and implemented in the real world.

Although the approach we have advocated above is similar to that devel-
oped by Besley and Coate, one would have to inquire into the origin of a new
instrument, if one adopts our theoretical perspective. In many cases, partic-
ular existing instruments may not be employed in all political equilibria. As
a result, it would not be unusual for shocks to lead to new equilibrium solu-
tions, where some previously unused instruments (already available before the
shock) become part of the winning political strategy. In such a formulation, the
model contains no truly exogenous instruments, and the Besley/Coate analysis
does not apply.

The approach of Besley and Coate seems appropriate, on the other hand, if
the appearance of a new instrument results from an exogenous shock created
by the appearance of a novel idea. Unless we can explain generation of the
idea, however, our analysis of such a case refers will refer primarily to the
effects created by the shock, rather than to the mechanism responsible for the
emergence of new policy instruments.

Because they also see all policy instruments (whether currently in use or
not), as being determined in an equilibrium, authors such as O'Flaherty and
Bhagwati (1997) recommend giving advice to citizens rather than to govern-
ments, in an attempt to influence public opinion, and thereby to move politi-
cal equilibrium in a desirable direction. However, as Carl Shoup pointed out
(1991, personal communication), "any policy framer who adopts an 'activist'
approach puts himself outside the scope of [the] welfare analysis, for he be-
comes then just one of the combatants in the struggle to get for himself the
most with the least pain." The general issue here concerns the place where
the policy analyst is to stand, once all political behavior and all instrument
use has been endogenized.[24] The answer, like that given in the literature on
contemporary industrial organization, may be found by studying the operation
and reform of institutions. Solutions found in this way will only be partially
satisfactory, however, since (as pointed out earlier in our discussion of the
Leviathan model) self-interested politicians must in the end decide themselves
to alter existing institutional or constitutional arrangements.

6.   CONCLUSION

Economists have developed a variety of rules or norms to deal with tax
analysis. In this chapter, we review eight such rules, dividing them into two
categories, depending on whether they focus primarily on outcomes or on
process.

As the discussion demonstrates, outcome-oriented tax norms are derived from analytical frameworks that do not acknowledge the need for collective choice. Unless the effects of the operation of collective choice mechanisms are explicitly recognized in the framework of analysis, however, we cannot tell if the policy proposals that emerge will be consistent with political equilibrium. Nor is it possible to determine in all cases whether such proposals will result in actual welfare improvements.

Our analysis also leads to several criticisms linked to particular norms. Lump sum taxes, often used as an outcome-oriented norm, are not likely to emerge as a viable structural solution in a democratic equilibrium. Neutrality, another such norm, if used as a guide to policy reform, will not be consistent with political competition that pushes governments to differentiate the tax treatment of different economic activities in the real world. Nor can we expect rules for international tax harmonization, based largely on the assumption that governments behave in a benevolent manner, to be robust in situations where competition among governments helps to constrain political opportunism.

Outcome-oriented rules attempt to overcome the extensive information problem associated with implementation of a tax blueprint by a planner with the help of simplified planning guidelines. However, an economically efficient tax system in a modern economy must of necessity be complicated, since a myriad of economic margins must be taken into account. Application of simplified rules requires many compromises that may reduce welfare. An alternative approach would be to decentralize the tax policy process. This would allow competitive political pressures, coupled with a decentralized and specialized bureaucracy, to generate information for decision makers, without anyone being aware of what happens in the system as a whole. Such decentralization is a classic solution to economic information problems associated with fiscal and other types of planning. Supporters of outcome-oriented, centrally applied rules must answer the question why the operation of a decentralized competitive political system would not generate results that are preferable to what can be achieved with outcome-oriented norms.[25]

The process-oriented literature also promotes simplified solutions. Examples include the broadly based income tax, flat taxation without exemptions, and the use of independent tax commissions. Unlike outcome-oriented rules, however, these proposals are intended as constraints on the nature of the political process. Those who advocate particular tax structures as political constraints believe that their proposals will improve the democratic process. It is not clear, however, why the proposed structures should be consistent with political equilibrium, and why they should be superior to tax structures generated by the workings of a competitive political process.

Our discussion suggests that a more comprehensive type of normative tax analysis is needed. Such an analysis would include a standard of reference,

against which actual outcomes are to be compared, that explicitly incorporates collective choice mechanisms. Until such a revised framework has been created, existing norms may continue to provide partial guidance. The useful insights that can be derived from currently accepted policy norms must be tempered, however, with a careful evaluation of the biases that arise in applying such rules to a functioning democratic process.

NOTES

1. It may be useful to point out that we do not examine ethical norms that provide a prescriptive guide for appropriate individual behavior.
2. It is possible to find a special tax scheme that will overcome the preference revelation problem under certain conditions, such as the Clark-Groves and Ledyard-Groves mechanisms (see Cornes and Sandler, 1996, pp. 221-234 for discussion and references). However, none of these schemes appear to be a practical method of financing a modern public sector.
3. The compulsory nature of taxation referred to here stems from the desire to overcome the problem of free-riding. Issues of coercion in a stronger sense are dealt with below. See Buchanan (1975) for further discussion of the distinction between the production of public goods, prevention of free-riding, for which compulsory taxation is usually required, and coercive redistribution in the sense of taking without compensation.
4. Individuals substitute less valuable, but less heavily taxed activities for activities that are taxed more highly, until further reductions in tax liabilities are just offset by the loss in welfare as a result of the substitution. Since the loss of revenue to the government is also a gain to the taxpayer, it is not lost to society as a whole. However, the welfare loss from the shift towards less valuable activities remains as a net loss (over and above that due to the payment of taxes) to the individual taxpayer and to society. It is important to note that the substitution of taxed activity that is the essential source of excess burdens would not exist if taxing and spending were not separated. For in that case, the substitution away from taxed activity would also lead directly to a corresponding and equal loss of valued public services. See Creedy (1998) for discussion of the definition and measurement of the excess burden of taxation, which is also often referred to as the deadweight cost of taxation.
5. Here we assume away the problem of actually identifying individual taxpayers in such a manner that it is not possible for them to avoid the tax. See for example, Jha (1998, pp. 294-295) and Myles (1995, pp. 44-48) or other public finance texts such as Cullis and Jones (1998), Rosen (1999) or Stiglitz (1988).
6. A large literature has arisen to deal with the problems of measuring excess burdens from various types of taxes. For a discussion of the issues, see, for example, Creedy (1998).
7. This particular result requires that all cross-price elasticities of demand be zero. See Jha (1998, p. 294) for details.
8. Stiglitz (1997) and others have suggested the use of the Pareto criterion in place of social welfare maximization. However, in our view, distributional goals remain essential in much of the literature.
9. Bird also draws attention to the "chasm" that exists between optimal tax theorists and practitioners (1991, 38).
10. Some authors refer to a harmonized tax system as a system in which different jurisdictions adopt the same set of tax rates. We employ the term in its more widely used sense to refer to a regime in which the tax systems of different countries, or of provinces or states within a federal system, are coordinated according to one set of general principles.

11. Wilson (1999), Oates (1999) and Musgrave (1991) survey the literature on interregional and international harmonization more fully.
12. See also Keen (1996).
13. To guard against a raid on the domestic treasury, foreign tax credits are usually limited to the domestic tax that would be payable if the foreign source income was earned at home.
14. Obviously the problem of smuggling arises here and must be dealt with in theory and practice.
15. *Tax Reform for Fairness, Simplicity and Economic Growth* (1984) and *the President's Tax Proposal* (1985), both issued by the U.S. Treasury, summarize the government's proposals and the main issues in the debate preceding the 1986 reforms.
16. See Kydland and Prescott (1977) for the seminal statement of this problem. Drazen (2000, Chaps. 4-6) provides a recent review of the literature.
17. The intuition behind the efficiency of the political equilibrium is the following: If the Pareto frontier has not been attained, it will be possible for some party to promise welfare gains to some without reducing the welfare of others, thereby increasing its chances of electoral success. Political competition will drive parties to seek out all such politically profitable and economically superior reforms.
18. For a recent exploration in this direction, see Hotte and Winer (2000).
19. Coate (2000) also advocates this sort of procedure.
20. For an interesting review of the historical debate, see Simon (1981, Chap. 2).
21. For further discussion of this approach, see Winer and Hettich (1999) and Hamilton and Slutsky (2000).
22. The literature is struggling to come to grips with the problem. Lockwood (1992) and Eggert (1999) show how the standard prescriptions for international harmonization can be overturned by altering the assumptions concerning the objectives pursued by any government. See also Edwards and Keen (1996) and Schulze and Ursprung (1999). For recent reviews of the literature, see Oates (1999) and Shaviro (2000).
23. For further discussion of the issues than we provide here, see Drazen (2000).
24. For contributions to this debate, see the November 1997 special issue of Economics and Politics (T. N. Srinivasin (co-ed). See also Pomery (1991) and Kirschgassner (1999) for interesting contributions to the debate on the nature of policy analysis when policy instruments are endogenized.
25. It is to our mind evidence of the incompleteness of the outcome-oriented approach that nowhere in this literature can one find any discussion of the potential benefits from increasing the extent of political competition.

REFERENCES

Atkinson, Tony and Stern, Nicholas. (1974). "Pigou, Taxation and Public Goods." *Review of Economic Studies* 41: 119-128.

Besley, Timothy and Coate, Stephen. (2002). "On The Public Choice Critique of Welfare Economics." *Public Choice,* forthcoming.

Bird, Richard. (1991). "Tax Administration and Tax Reform: Reflections and Experience." In *Tax Policy in Developing Countries,* Khalilzadeh-Shirazi, Javad and Shaw, Anwar, eds., pp. 38-56. The World Bank.

Boylan, Richard T. and McKelvey, Richard D. (1995). "Voting Over Economic Plans." *American Economic Review* 85: 860-871.

Brennan, Geoffrey and Buchanan, James M. (1980). *The Power to Tax.* New York: Cambridge University Press.

Buchanan, James M. (1975). *The Limits of Liberty: Between Anarchy and Leviathan.* Chicago: University of Chicago Press.

Coate, Stephen. (2000). "An Efficiency Approach to the Evaluation of Policy Changes." *Economic Journal* 110 (463): 437-455.

Cornes, Richard and Sandler, Todd. (1996). *The Theory of Externalities, Public Goods and Club Goods,* 2nd edition. New York: Cambridge University Press.

Creedy, John. (1998). *Measuring Welfare Changes and Tax Burdens.* Cheltenham UK: Edward Elgar.

Cullis, John and Jones, Phillip. (1998). *Public Finance and Public Choice,* 2nd edition. Oxford University Press.

Drazen, Allan, (2000). *Political Economy in Macroeconomics.* Princeton, N.J.: Princeton University Press.

Edwards, J. and Keen, M. (1996). "Tax Competition and Leviathan." *European Economic Review* 40: 113-134.

Eggert, Wolfgang. (1999). "Capital Tax Competition with Inefficient Government Spending." Paper given at the IIPF congress, Moscow, August 1999.

Foley, Duncan. (1977). "Resource Allocation and the Public Sector." *Yale Economic Essays* 7 (1): 45-98.

Fischer, Stanley. (1980), "Dynamic Inconsistency, Cooperation and the Benevolent Dissembling Government." *Journal of Economic Dynamics and Control* 2: 93-107.

Gillis, Malcolm. (1989). "Tax reform: Lessons from Postwar Experience in Developing Nations," In *Tax Reform in Developing Countries,* Gillis, M., ed. Duke University Press.

Hamilton, Jonathan and Slutsky, Steven. (2000). "Decentralizing Allocation and Distribution by Separation with Information Transfers." *Journal of Public Economic Theory* 2 (3): 289-318.

Hayek, F. A. (1945). "The Use Of Knowledge in Society." *American Economic Review* 35 (4), September.

Hettich, Walter and Winer, Stanley L. (1999). *Democratic Choice and Taxation: A Theoretical and Empirical Analysis.* New York: Cambridge University Press.

Hettich, Walter and Winer, Stanley L. (1997). "The Political Economy of Taxation." In *Perspectives on Public Choice,* Mueller, D., ed., pp. 481-505. New York, Cambridge University Press.

Jha, Raghbendra. (1998). *Modern Public Economics.* New York: Routledge.

Keen, Michael. (1996). "The Welfare Economics of Tax Coordination in the European Community." In *The Economics of Tax Policy,* Devereux, Michael P., ed., pp. 189-214. Oxford University Press.

Kirschgassner, Gebhard. (1999). "On the Political Economy of Economic Policy Advice." In *The Transfer of Economic Knowledge,* Mohr, Ernst, ed., pp. 13-31. Cheltenham: Edward Elgar.

Kotlikoff, L., Persson, T., and Svensson, L. (1988). "Social Contracts as Assets: A Possible Solution to the Time Inconsistency Problem." *American Economic Review* 78: 662-677.

Lindahl, Eric. (1919). *Die Gerechtigkeit der Besteuerung.* Reprinted in part as "Just Taxation: A Positive Solution" (translation by Elizabeth Henderson). In *Classics in the Theory of Public Finance,* Musgrave, R. A. and Peacock, A. T., eds. London: Macmillan, 1958.

Lockwood, B. (1992). "Commodity Tax Competition Under Destination and Origin Principles." *Journal of Public Economics* 52: 141-162.

Lohman, Susan. (1998). "Reputational versus Institutional Solutions to the Time Consistency Problem in Monetary Policy." In *Positive Political Economy: Theory and Evidence,* Eiffinger, S. and Huizinga, Harry P., eds., pp. 9-28. Cambridge: Cambridge University Press.

Mirrlees, James A. (1971). "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38: 175-208.

Musgrave, Peggy. (1991). "Fiscal Coordination and Competition in an International Setting." In *Retrospectives on Public Finance,* Eden, L., ed., pp. 276-305. Durham: Duke University Press.

Myles, Gareth D. (1995). *Public Economics.* Cambridge U.K.: Cambridge University Press.

Oates, Wallace E. (1999). "Fiscal Competition and European Union: Contrasting Perspectives." Unpublished paper presented at the conference on Fiscal Competition and Federalism in Europe, Mannheim, Germany, June.

Persson, Torsten and Tabellini, Guido. (1990). *Macroeconomic Policy, Credibility and Politics.* London: Harwood.

Pomery, John. (1991). "The New Trade Theory: Implications for Policy Analysis." In *Policy Analysis and Economics: Developments, Tensions, Prospects,* Weimer, David, ed., pp. 159-186. Kluwer Academic Publishers.

Ramsey, F. P. (1927). "A Contribution to the Theory of Taxation." *Economic Journal* 37: 47-61.

Rosen, Harvey S. (1999). *Public Finance,* 5th edition. Boston: Irwin/McGraw-Hill.

Royal Commission on Taxation. (1966). Chaired by Kenneth LeM. Carter. Report. Ottawa: Queen's Printer (The Carter Report).

Schulze, G. G. and Ursprung, Heinrich. (1999). "Globalization of the Economy and the Nation State." *World Economy* 22: 295-352.

Schumpeter, Joseph (1954). *History of Economic Analysis.* Oxford: Oxford University Press.

Seligman, Edwin R. A. (1921). *Essays in Taxation,* 9th edition. New York: Macmillan Co.

Shaviro, Daniel. (2000). "Some Observations Concerning Multi-Jurisdictional Tax Competition." Forthcoming in *Regulatory Competition and Economic Integration: Comparative Perspectives,* Esty, Daniel and Geradin, Damien, eds, Oxford University Press.

Sheffrin, Steven M. (1989). *The Making of Economic Policy: History, Theory, Politics.* Cambridge, MA: Basil Blackwell.

Simon, Herbert A. (1981). *The Sciences of the Artificial,* 2nd edition. MIT Press.

Simons, Henry. (1936). "Rules Versus Authorities in Monetary Policy." *Journal of Political Economy* 64(1): 1-30.

Simons, Henry. (1938). *Personal Income Taxation: The Definition of Income as a Problem of Fiscal Policy.* Chicago: University of Chicago Press.

Stiglitz, Joseph E. (1997). "Pareto Efficient and Optimal Taxation and the New New Welfare Economics." In *Handbook of Public Economics, Vol II,* Auerbach and Feldstein, M., eds., pp. 992-1042. North-Holland.

Stiglitz, Joseph E. (1988). *Public Sector Economics,* 2nd edition. New York: Norton.

U.S. Department of the Treasury (1984). *Tax Reform for Fairness, Simplicity and Economic Growth 3.* Washington, DC: U.S. Government Printing Office.

U.S. Department of the Treasury (1985). *The President's Tax Proposal.* Washington, DC: U.S. Government Printing Office.

Wilson, John. (1999). "Theories of Tax Competition." *National Tax Journal* 52 (2): 269-304.

Winer, Stanley L. and Hettich, Walter. (1999). "Information, Coordination and Tax Policy Making." In *Rationality on Public Policy: Retrospect and Prospect, A Tribute to Douglas G. Hartle,* Bird, R., Trebilcock, M., and Wilson, T., eds., pp. 19-42. Toronto: Canadian Tax Foundation.

Chapter 6

TAXATION, PRODUCTION, AND REDISTRIBUTION

Randall G. Holcombe
Florida State University
holcombe@garnet.acns.fsu.edu

Abstract   In one sense the public choice revolution of the 20th century ocurred because economic and political analysis now routinely takes into account the incentives of political decision-makers, and recognizes that political decision-making can result in inefficiencies. In another sense, this public choice revolution is seriously incomplete, because while it has succeeded in producing a theory of government failure, it has not taken the next step to develop a framework for optimal public policy in light of the characteristics of collective decision-making. This chapter takes a step in that direction by sketching out optimal policies for taxation, production and redistribution, taking into account that these public policies are products of the political system. Public choice might be thought of as applying the tools of economics to analyze political decision-making. This chapter uses the results of public choice to redevelop the theory of public finance.

At the beginning of the 21 st century, few economists would take issue with the idea that to maximize the well-being of a nation's citizens, market allocation of resources is superior to government planning.[1] At the same time, most economists would also argue that some role for government is necessary in an economy to undertake those activities for which the market is not well-suited.[2] Some of those activities might be fundamental requirements for the operation of a market economy. For example, market exchange is predicated on the clear definition and protection of property rights, and the protection of

property rights is one of the roles that has traditionally been undertaken by government.[3] Other government actions might be desirable if they could overcome problems that markets might have in allocating resources, such as may arise with externalities and public goods. The purpose of this chapter is not to debate the optimal role and scope of government, but rather to look at the methods by which economic analysis has dealt with those questions.

The conventional method of public finance has been to identify problems that might arise in the market allocation of resources and then to determine some type of optimal policy that can be used to correct this problem. The typical analysis assumes that the government can costlessly identify and implement this optimal policy. One of the important insights of public choice, as Buchanan (1975) clearly explains, is that there may be inefficiencies in government allocation of resources too, and that methodologically, an analysis should treat public sector decision-makers in the same framework as private sector decision-makers. Government can fail to allocate resources efficiently for many reasons, ranging from having inadequate information to understand or solve the problem to being unable to implement or enforce the optimal solution. This chapter ignores many of these very real problems, and focuses solely on the problems that arise from the collective decision-making process. The chapter unrealistically assumes that a benevolent dictator would be able to produce a government policy that would result in an optimal use of resources, in order to focus on the problems that arise solely as a result of having to decide on what policy to follow through democratic decision-making.

For example, one might see that air pollution is caused by excessive use of an unpriced resource, which creates an externality. However, internalizing the externality might require more information than the government has regarding how much pollution is coming from various sources, and what regulations and/or pricing mechanisms would result in the optimal control of the externality. Even if the government could figure out the optimal course of action, it may not be able to force citizens to follow the optimal course of action it has calculated. These problems are assumed away, however, to focus on the problems that arise solely as a result of the government's decision-making process.

Public choice theory already has much to say about the inefficiencies of the collective decision-making process. The purpose of this chapter is to apply that body of public choice theory to the issues dealt with by standard public finance. There is a well-established theory of optimal taxation, for example, but that theory does not take into account the fact that taxes are the product of a political decision-making process. This chapter looks at optimal taxation from a public choice perspective and concludes that an optimal tax system is significantly different if it is produced by a democratic decision-making process than if it is created by a benevolent dictator. Similarly, optimal public sector production and redistribution policies are different when they are generated

through collective decision-making rather than imposed by a benevolent dictator. Optimal taxation, production, and redistribution policies are so different when one takes into account the fact that policies are produced as a result of a collective decision-making process that this chapter suggests a wholesale rethinking of the basic principles of public finance, based on a public choice analysis.

In one sense, a public choice revolution occurred in the last half of the 20th century, because economic and political analysis now routinely takes into account the incentives of political decision-makers, and recognizes that political decision-making can result in inefficiencies. In another sense, this public choice revolution is seriously incomplete, because while it has succeeded in producing a theory of government failure, it has not taken the next step to develop a framework for optimal public policy in light of the characteristics of collective decision-making. This chapter takes a step in that direction by sketching out optimal policies for taxation, production and redistribution, taking into account that these public policies are products of the political system. Public choice might be thought of as applying the tools of economics to analyze political decision-making. This chapter uses the results of public choice to redevelop the theory of public finance.

1.   OPTIMAL TAXATION

The concept of optimal taxation has been long established in public finance, going back at least to Ramsey (1927), who laid out the conditions for an optimal schedule of excise taxes. Ramsey's idea was that an optimal tax system was one that was designed to minimize the excess burden of taxation. More recently, Mirrlees (1971, 1976) has extended optimal tax theory in several ways, but retains Ramsey's original idea that optimal taxes minimize the welfare loss from taxation. While Ramsey considers only the deadweight loss from taxation, Mirrlees is willing to make interpersonal utility comparisons to weigh the relative losses of one person against another. Still, Ramsey's fundamental concept remains intact. An optimal tax system is one that minimizes the excess burden of taxation for the collection of a given amount of revenue.

A more complete analysis of the welfare costs of taxation would include compliance and enforcement costs (Slemrod and Sorum, 1984; Slemrod, 1990). Not only does the tax system impose an excess burden on the economy, it also imposes direct costs on both the government and taxpayers. Government expends resources to collect taxes, monitor payments, and enforce the law, and taxpayers incur costs from keeping records, tendering payment, and responding to government enforcement inquiries. Thus, minimizing the welfare loss of taxation must include not only the excess burden of taxation, narrowly defined, but also the cost of the resources used by both government and taxpayers in the collection process.

Even adding compliance and enforcement costs to the deadweight loss of taxation gives an incomplete measure of the welfare cost of taxation, however. The standard public finance analysis of taxation assumes that the tax system is created costlessly, rather than through a political decision-making process. A full accounting of the welfare costs of taxation must include the cost of the political process that produces the tax system. Thus, a complete taxonomy of the welfare costs of taxation would include the deadweight loss of taxation, compliance costs (borne by taxpayers), enforcement costs (borne by government), and political costs. The political costs of taxation arise from several sources. Most obvious are the costs incurred in the collective decision-making process—the decision-making costs discussed by Buchanan and Tullock (1962). However, the rent-seeking costs identified by Tullock (1967) and Kreuger (1974) are undoubtedly more significant.

Rent seeking costs are generated because people want to influence the tax system for their benefit. Taxpayers engage in political activity to try to get their taxes reduced. Thus, any existing tax will generate political costs incurred both by those taxpayers who oppose the tax and by supporters—who perhaps are beneficiaries of the spending the tax finances—who want to keep the tax in place. A tax does not even have to exist to generate political costs. New taxes (and increases in the rates of existing taxes) are being proposed continually, so political costs are incurred by potential taxpayers who must always be on their guard against new taxes.

Estimating the magnitude of the political costs of the tax system would be a difficult task. Holcombe (1997a, b) presents some data to suggest that the political cost of a selective excise tax exceeds ten percent of the revenues raised by the tax. Political costs in this range would exceed the magnitude for enforcement costs and compliance costs combined, suggesting the importance of incorporating political costs into the economic analysis of taxation. When one reflects on the political process behind the tax system, such an estimate does not appear unreasonable, however. When new taxes are being debated, both proponents and opponents engage in political activity to support their positions, and after taxes are passed taxpayers lobby to try to get them eliminated or reduced. Political costs may be even more subtle, as potential taxpayers maintain a lobbying presence with legislators to try to prevent legislation opposed to their interests from ever surfacing. Thus, the nature of the system means that political costs are being generated even when there is no revenue produced, as a result of political activity of those who want to keep from being taxed. Because the tax system is a product of the political system, and taxes can be modified through political action, taxpayers must participate in the political process in order to protect themselves from being the victim of political predation from those who want to use the political system to transfer resources from some toward others.

A fundamental conclusion from the public choice revolution is that public policy is produced by the actions of individuals within a political process, and because of the incentive structure of the public sector, political processes will not always allocate resources efficiently. Applying this idea to the concept of optimal taxation implies that political costs must be included as a part of the welfare cost of taxation. Thus, an optimal tax system will minimize the sum of the deadweight loss of taxation, enforcement costs, compliance costs, and political costs. Political costs have been barely recognized in the public finance literature, and while there are not good estimates of the magnitude of political costs, once they are recognized it is apparent that they are a significant part of the total cost of the tax system.

Two related implications follow. The first, following Buchanan (1975), is that the tax system that is produced by the political process is likely to deviate substantially from what tax theorists would argue is optimal. The second is that from a policy perspective, actually designing an optimal tax structure means designing one that is insulted from political pressures, which therefore can reduce political costs. Not everybody would agree with these implications. For example, Hettich and Winer (1988) and Winer and Hettich (1998) argue that political incentives tend to create an efficient tax system, following the ideas of Becker (1983) and Wittman (1989). These counterarguments will be examined toward the end of this chapter, to evaluate how well the arguments here hold up to criticism. Meanwhile, if political costs really are a significant component of the welfare loss of taxation, consider how these costs might be minimized.

2.   REDUCING THE POLITICAL COSTS OF TAXATION

The political costs of taxation arise because people use resources to try to modify the tax system for their benefit. Thus, the political costs of taxation can be reduced by creating a tax system that is relatively unchangeable, and can be modified only when modifications are supported by a broad consensus of opinion. A tax system that is not responsive to rent-seeking activity by interest groups will lower the amount of rent seeking, thereby reducing political costs. Thus, Buchanan (1967) calls for the creation of a fiscal constitution that dictates the basic tax structure, and allows change only if there is a substantial consensus. Following the distinction made by Buchanan and Tullock (1962), the tax structure should be a part of the constitution, rather than a product of post-constitutional decision-making. The creation of a fiscal constitution that could be modified only with substantial agreement is that the payoff from lobbying to try to change the tax structure would fall, which would reduce rent-seeking and lower the political costs associated with the tax system.

The traditional concept of optimal taxation implies that the tax system is fine-tuned to reduce the deadweight loss of taxation. Once political costs are incorporated into the analysis, opportunities to fine-tune the tax system for efficiency also open opportunities for rent-seeking to modify the tax system. Thus, rather than designing the tax system so it can be adjusted easily to conform to marginal conditions for optimality, as in the traditional public finance theory of optimal taxation, once political costs are taken into account, the tax system should be designed so that it can retain political support in the face of changing conditions, and so that modifications can be undertaken in such a way as to promote consensus rather than rent-seeking competition. If the tax system can be modified easily to minimize the deadweight loss of taxation, it will also be open to rent-seeking by special interests who want to modify the tax system for their own benefit. These principles can be illustrated by applying them to commodity taxation and income taxation, and contrasting the implications with those of the traditional theory of taxation.

3.   COMMODITY TAXES

Optimal commodity taxation, following Ramsey (1927) implies setting different tax rates for different commodities such that when goods are produced in competitive conditions, the ratio of tax rates for commodities is set at the inverse of the ratios of their elasticities of demand. In theory the Ramsey rule minimizes the excess burden of commodity taxation, but it does not take into account administrative, compliance, and especially political costs. The practical problem with applying the Ramsey rule is that there is no market-generated indicator of elasticities of demand for commodities. They have to be estimated, and they can change over time. An omniscient observer applying the Ramsey rule could simply observe demand elasticities and set tax prices in inverse relation to them, changing the taxes whenever demand elasticities changed. In a political setting, somebody has to provide estimates of demand elasticities so that the Ramsey rule can be applied. Producers of taxed goods have an incentive to employ experts, both economists and lobbyists, to argue that the demand for their output is more elastic than other taxed goods, pitting the producers of taxed goods against each other in a rent-seeking battle to raise the government's official estimated elasticity of demand for their goods. Applying different tax rates to different goods, especially when those tax rates are based on difficult-to-observe criteria, invite rent-seeking and political costs. Once political costs are taken into account, the Ramsey rule is no longer optimal.

To minimize the political costs of excise taxation, the relative tax rates among goods should be constitutionally fixed. The simplest method of doing

this would be to have the same rate apply to each good. If the fiscal constitution specifies that all excise taxes will be levied at the same rate, the possibility of rent-seeking to get a lower rate, or avoid a higher one, is eliminated, reducing political costs. Under this regime, any argument for increased excise tax revenues would imply raising all excise tax rates, which would increase the likelihood of consensus. If the consensus was that more public sector revenue really is desirable, then there would be general support for an increase in excise taxes, whereas any lobbying for lower excise tax rates would have to be aimed at lower rates for everybody, rather than for any one narrow interest. Applying these public choice principles to excise taxation provides a good example to show how optimal tax theory is changed when political costs are taken into account. Optimal excise taxation in traditional tax theory follows the Ramsey rule that goods should have tax rates set in inverse proportion to their elasticities of demand. When political costs are taken into account, a fiscal constitution that mandates that all goods are taxed at the same rate is more likely to be optimal.

In order to determine if a single rate for all goods is really preferred to differential rates, one would have to begin by comparing the political costs associated with differential rates against the reduction in excess burden, if any, that came with allowing differential rates. If the political costs were small relative to the excess burden that could be relieved with differential rates, this might point toward applying the Ramsey rule. Even here, however, some caveats are required. Changing conditions could change the relative sizes of the political costs and excess burden, and at that point, enacting a constitutional change might prove too costly to implement. Thus, the argument tips toward favoring the same excise tax rate for all goods regardless of the relative costs.

An even stronger argument against trying to implement the Ramsey rule is that because tax rates are determined by the political process rather than by an objective measure of relative elasticities of demand, the political system would be unable to produce tax rates that satisfy the Ramsey rule anyway. If the political objective were to set excise tax rates according to the Ramsey rule, in reality tax rates would be determined at least as much by the political power of different groups of taxpayers as by the elasticities of demand of the goods they were selling. Even if the Ramsey rule is used as a benchmark, when considering the politics involved in setting tax rates, it is not unlikely that setting all excise taxes to the same rate would result in a lower deadweight loss than using the political process to try to produce rates that satisfy the Ramsey rule. When considering all factors, there is a strong case to be made that when political costs are taken into account, optimal excise taxation implies that the same rate be applied to all goods, rather than trying to set rates in inverse proportion to the elasticities of demand.

4.   INCOME TAXATION

A similar case can be made for proportional income taxation. Under progressive income taxation, there is no clear guide as to how progressive rates should be. This encourages rent-seeking and political costs as those at the bottom end of the income distribution argue for more progressive rates while those at the top end expend resources to try to make rates less progressive. As Hayek (1960, p. 313) notes, only proportional income taxation brings with it a principle which, once adopted, is relatively easy to defend. As a policy, proportional taxation seems to have much to support it. The optimal tax literature following Mirrlees (1971, 1976) makes interpersonal utility comparisons to use the tax system as a redistributive tool to maximize social welfare. One might object to this line of analysis on methodological grounds, but even using this very different methodology, the optimal tax literature finds that proportional income taxation is very close to optimal. By making proportional rates a part of the fiscal constitution, rent-seeking over rates is eliminated and political costs are minimized. Any argument for higher rates means higher rates for everybody, minimizing conflicts of interest.

In the case of personal income taxation, the incorporation of political costs into the analysis reinforces the conclusion of the optimal tax literature which—using very different methodology—supports proportional income tax rates. In the case of excise taxation, the incorporation of political costs suggests uniform excise tax rates rather than rates differentiated by demand elasticities, in contrast to traditional public finance; in the case of income taxation, the incorporation of political costs reinforces the traditional literature's conclusion in favor of proportional income taxation. In both cases, the incorporation of political costs adds considerable insight to the analysis.

5.   THE BENEFIT PRINCIPLE

Up to this point, the analysis of taxes has been undertaken without considering how the tax revenue would be spent, but as Buchanan (1976) notes, taxes are the price we pay for government goods and services, so any determination of an optimal tax structure must take into account how the revenues are spent. This implies that the benefit principle should be embodied as a part of the fiscal constitution, and taxes should be paid in proportion to the benefits received from public sector output. User charges should be applied whenever feasible, and in any case taxes should be levied on those who benefit from the output they finance. This idea is not new. It is the foundation supporting the well-known ideas of Lindahl (1919) and Wicksell (1896), who take an approach that recognizes the political process within which taxes are approved, much as is done in this chapter.

By establishing the benefit principle as a part of the fiscal constitution, both the political costs of taxation and the excess burden can be minimized. The incentive to rent-seek is minimized because people pay taxes in proportion to the benefits they receive. This insulates people from being forced through the coercive power of government to pay taxes to finance benefits for others. Furthermore, because taxes levied according to the benefit principle have more of the characteristics of market prices, the excess burden will be lower. Market prices have no excess burden because they serve to ration goods and allocate them to those who value them the most. Taxes levied according to the benefit principle could serve the same role. People would lobby for goods only if they also were willing to pay for them. The benefit principle, which connects the payment of taxes with the production of public sector output, points directly toward a more detailed analysis of public production.

6.   PRODUCTION

The standard economic theory of public sector production rests on the concept of market failure. Following this theory, perfectly competitive markets allocate resources efficiently, but when markets are imperfect or not competitive, resources are misallocated. Following the traditional analysis, which is clearly laid out by Bator (1958), when a market failure occurs, government can correct the market failure through appropriate policy. Samuelson's (1954, 1955) public goods theory has been particularly influential along these lines, arguing that when goods exhibit jointness in consumption, the market will fail to provide the optimal amount, so government production is required for optimality. This market failure theory of government production can be challenged on many grounds,[4] but perhaps the most substantial criticism is that it does not take account of the political process that determines what goods the public sector produces.

There is a well-developed interest group theory of public sector production which suggests that government production is the outcome of a political process that is guided by the economic interests of those involved, and that the resulting government output is not likely to be optimal using the benchmarks of neoclassical welfare economics. Buchanan (1975) argued that public sector production should be modeled in the same way as private sector production, taking into account the incentives of all participants in the process, and, one should add, the information available to them. Analysis done in this way leads to an interest group model of government (Weingast, Shepsle and Johnsen, 1981; Holcombe, 1985) in which government production caters to narrow special interests because nobody has an incentive to promote the general public interest.

The market failure theory of government production violates the very tenets of neoclassical economics upon which it is built. According to the neoclassical

theory, markets fail because people respond to market incentives, acting in their own self-interest. Yet the neoclassical solution is to turn over production of public goods to the government, where supposedly people act in the public interest. A more consistent theory would explain government production as a result of the self-interested actions of those in government. An example can illustrate how a theory of public goods based on rational self-interest can be developed.

7.   SELF-INTEREST AND PUBLIC PRODUCTION

Why does government produce national defense? The traditional answer is that national defense is a public good, and will be underproduced in the private sector because of the free rider problem, causing a market failure. Thus, the government steps in, acting in the public interest, to produce the optimal amount and remedy the market failure. Note that this explanation requires the government to act in the public interest. An alternative explanation, developed in more detail by Holcombe (1994, b), is that the government gets its income from taxing the productivity of its citizens. Thus, it has an incentive to protect their productive assets, because in doing so it is protecting its source of income. Seeing national defense in this way provides the foundation for an exchange model of government, where government trades protection for tribute. Citizens want to be protected, so they accept the payment of their tax dollars in exchange for national defense, while the government wants the tribute, and wants to protect its source of income, so it willingly produces national defense in exchange for its tax revenues.

This exchange theory of national defense is more consistent with basic economic principles than the market failure theory. The exchange theory relies on the self-interest of the parties involved, whereas the market failure theory requires altruistic behavior on the part of those in government. From this foundation, an entire theory of government can be built, relying on the self interest of all parties. This basic model of national defense, explained in more detail in Holcombe (1994), depicts government production as a result of the self-interested actions of those who run government. Government has an incentive to protect its source of income, which is tax revenue levied on the productive capacity of the private sector. Other government production, such as the provision of postal services, utilities, and a host of other goods, are the result of government using its coercive capacity to give itself a monopoly, and reap monopoly profits. Government has an incentive to provide public education to propagandize its citizens and control the flow of information, as Holcombe (b) explains. Some government production is undertaken to provide benefits to interest groups that are in a position to reciprocate by providing support to the government. One can analyze individual government programs to see that

regardless of what public interest justifications one can give for government production, government production always benefits those in government.

Looked at in this way, government production is the result of government using its coercive abilities to reap monopoly profits, or to trade production for the benefit of some interest group in exchange for political support, rather than to correct market failures. Economic models in general are based on the depiction of individual behavior as self-interested. The actions of those in government can be modeled the same way, as Buchanan (1975) recommends. The result is entirely consistent with the observed activities of government.

The market failure theory of government production has another problem. As Samuelson (1954) notes, public goods present a revealed preference problem, in that because of free riders there is no way to discover the true demands of consumers for public goods. This problem extends more generally to any public sector production, because people always have an incentive to say they want to pay less in taxes, and receive more in benefits. The information to optimally produce public goods is not produced by the market, and is not otherwise available to government.[5] Thus, even a government acting completely in the public interest would not have sufficient information at its disposal to actually correct market failures and allocate resources Pareto optimally. This same problem was observed when discussing optimal taxation. While an omniscient observer might be able to see the optimal allocation of resources, the information necessary to allocate resources optimally is not available to government policy makers, regardless of their motivations.

8.   PUBLIC PRODUCTION AND THE FISCAL CONSTITUTION

This line of analysis points toward a theory of public sector production that parallels the theory of taxation developed above. The market failure model of government production models government as acting in the public interest to allocate resources Pareto optimally, just as the optimal tax model has the government designing a tax system to minimize welfare losses. Both of the traditional models depict government acting solely in the public interest, and depict government using information that is not generally available to any economic actors in order to achieve its goals. Once the political process is explicitly incorporated into the model, one can see that those in government do not have the incentive to allocate resources optimally, and do not have the information to do so even if the incentives were right. Rather, government production takes place for the benefit of those in power, either to enhance their incomes directly or to provide benefits to interest groups that can provide benefits in exchange. The availability of benefits to interest groups encourages rent-seeking and the associated political costs.

When the market failure theory of government production is analyzed using the same criteria that underlie neoclassical economics, it falls apart. Government production cannot correct market failures because public sector decision-makers have no incentive to do so, and because even if they wanted to, they have insufficient information to do so. However, designing government so that it can respond to market failures produces institutions that encourage rent-seeking and political costs. From a policy standpoint, optimal government production should be treated the same way as optimal taxation, with a fiscal constitution that strictly limits the activities of government. By so doing, rent-seeking becomes less worthwhile, because government is constitutionally constrained from responding to it, so political costs are reduced. This sounds much like the concept of enumerated powers that is in the Constitution of the United States. The problem is that over the centuries, government has found a way to expand beyond the limits implied by a literal reading of the Constitution.

Consider the expanded role of government powers within the context of theories of government production discussed here. The market failure theory of government production would suggest that government should have the discretion to act in order to remedy any market failures that occur, so this market failure theory supports a broader and more discretionary role for government on the grounds of economic efficiency. In contrast, once the political decision-making process is incorporated into the theory, a government with constitutionally limited scope and powers appears more desirable. This shows how significantly policy implications can be affected when the political process is incorporated into economic analysis.

9.   REDISTRIBUTION

At the beginning of the 21st century the largest single activity undertaken by government is redistribution. A straightforward explanation of this fact follows from the preceding analysis of government production. When the scope of government is left poorly defined, allowing it to respond to whatever issues it sees fit, rent-seeking activities will lead government to produce benefits for special interest groups rather than activities in the general public interest, and the most obvious way to provide benefits to an interest group is through a straightforward transfer of resources. Redistribution is readily understandable within the public choice framework that this chapter is following, along with the conclusion that it leads to rent-seeking and the attendant political costs. Still, the redistributive role of government cannot be dismissed out of hand, because it is a fundamental component of the traditional economic view of the role of the public sector.

Musgrave (1959), in his classic treatise on public finance, lists redistribution as one of the three major functions of government (along with production

and stabilization). In the optimal tax literature pioneered by Mirrlees (1971, 1976), taxation plays a redistributive role to enhance social welfare. Hochman and Rodgers (1969) note that redistribution has the characteristics of a public good, so the private sector will not be able to engage in optimal redistribution, thus justifying the government's role. Not only is there a strong tradition in public finance theory supporting government redistribution, equity and efficiency have always been considered together as policy goals. Even Adam Smith (1776, p. 777) considered the benefit principle and the ability to pay principle as two pillars of the tax system. However, within the framework set out here, the issue is not the desirability of redistribution, but its feasibility using political institutions. As with taxation and public sector production, those in the public sector do not have the incentive to produce optimal redistribution, and even if the incentives were there, they do not have sufficient information to produce optimal redistribution. However, the institutions that allow government to engage in redistribution encourage rent-seeking and associated political costs.

10.   THE POLITICS OF REDISTRIBUTION

Stigler (1970) relates Director's law of income redistribution, which argues that redistribution flows from those who have economic resources toward those who have political power. Following this line of reasoning, redistribution is unlikely to offer much assistance to those who are really in need, because the needy have relatively little political power. Indeed, because groups that have economic resources also tend to have political power, much redistribution will take resources from the same people who receive it back in the form of redistribution. The largest redistributive programs in the United States are social security and Medicaid, which provide benefits to elderly citizens regardless of need. The taxes to finance this redistribution come from workers, conforming to Director's law of redistribution from those who have income that can be targeted, to those who have political power. Conventional wisdom is that social security benefits are secure because of the political clout of the elderly, again reinforcing Director's law. Redistributive programs such as farm price supports and college financial aid tend to benefit upper-income people, and subsidization of art and music, while a small part of total redistribution, is clearly aimed at the tastes of the well-to-do rather than the needy. If one takes a look at the actual nature of redistribution, it is apparent that it conforms more with the interest group model of government than the Hochman and Rodgers vision.

Some redistribution does come to the truly needy, partly because people with political power sympathize with them and lobby for benefits, but probably more so because if the poor viewed themselves as getting little in the form of

benefits from government, they might create civil unrest. The political science and public choice literature often notes that lower income is correlated with lower voter turnout, but rarely notes that lower income is also correlated with civil disturbances from protesting specific government acts (e.g., police brutality, court decisions) to reacting against general economic conditions through actions such as organized riots and looting to individual criminal activities. Often, economic benefits to the poor are justified as a way of making them more law-abiding, and thus reducing the threat they pose to those better off. Some redistribution can be seen as nothing more than a bribe to keep those at the bottom end of the economic spectrum from upsetting the system.

Furthermore, as Brennan and Lomasky (1993) note, the non-poor may cast their votes in favor of redistribution because it costs them little to do so. Voters know that their one vote will not affect the outcome of an election, so they can vote expressively, casting a vote for redistributive programs, or candidates who advocate them, for example, to make them feel more charitable. Their one vote will not alter the outcome of the election, but may make them feel better for their charitable impulses. Even if people were charitably inclined, it may make more sense for them to vote in favor of governmental redistribution, where the bulk of the cost will be paid by others, than to engage in charitable giving of their own, where any giving costs them directly. Analyzing the political process, some redistribution to the poor is entirely consistent with self-interested behavior, and with Director's law. Even so, the bulk of the benefits from redistribution does not go to the needy, again in line with Director's law.

As governments became more heavily involved in redistribution, one question that frequently arose is why any redistribution would occur in kind rather than in the form of monetary transfers. Friedman (1962) argues the efficiency of cash transfers over transfers in kind, but a number of writers (Akerlof, 1978; Blackorby and Donaldson, 1988; Bruce and Waldman, 1991) have argued that in-kind transfers can be more efficient than cash, largely because they can be directed at particular recipients more accurately. In a world where an omniscient benevolent despot runs the government, such targeting of redistributive benefits could, in theory, be efficiency-enhancing, but in the real world where transfers are determined by the political process, and where those who have more political power have greater ability to target transfers to themselves, redistribution in kind simply provides a way of making sure that transfers only go to those favored by the political elite. Many interest groups could compete for cash transfers, but if legislation specifies farm price supports, only farmers are eligible. If legislation specifies small business loans, then only small business owners are eligible. Transfers in-kind are a way of targeting political benefits to specific recipients and excluding others.

11. REDISTRIBUTION AND THE FISCAL CONSTITUTION

If redistribution programs could be incorporated into the fiscal constitution, then the rent-seeking and political costs associated with them could be sharply reduced, but the very nature of redistributive programs prevents them from being a part of the fiscal constitution. Recall that when discussing taxation, Hayek (1960, p. 313) argued that proportional taxation provides a principle that can be defended, thus limiting rent-seeking by groups who want to alter their relative tax shares. A tax increase for one means an increase for all; conversely, a cut for one means a cut for all. Progressive taxation, on the other hand, always leaves the degree of progression open to rent-seeking, because there is no principle to dictate how progressive a tax structure should be. Redistribution shares this characteristic with progressive taxation. There is no good way to write a constitutional rule to guarantee a transfer to some without opening up the system to rent-seeking by others. One can defend a principle that the people who receive government benefits should pay for them, but how would one argue on principle that the government should tax group A to pay for benefits to group B, and yet be prevented from taxing group C to provide benefits to group D?

Originally, the Constitution of the United States was viewed as preventing some from being taxed to provide benefits for others, but over time the Constitution was reinterpreted so that the fiscal constitution allowed redistribution, and following Musgrave (1959), even encouraged it. Higgs (1987) describes how the Progressive Era at the beginning of the twentieth century brought with it a change in ideology. In the nineteenth century, Higgs argues, American ideology viewed the role of government as the protection of individual rights, whereas in the twentieth century people thought that the government should protect their economic well-being in addition to just securing their rights. Similarly, Anderson and Hill (1980) note the role of the courts in transforming American government in the same way. This transformation toward a redistributive state can be clearly understood within the context of the earlier discussion on taxation and government production.

Even a government limited to protecting individual rights must engage in production or police and military services, and taxation to pay for those services. As described earlier, taxation and production lead to rent-seeking as people try to shift costs to others and shift benefits toward themselves. Eventually, such activities become so obviously redistributive that politically, it is a small step to ask for a simple transfer rather than a government benefit through some expenditure program or tax cut. Holcombe (1992) provides evidence that this transformation began before the Civil War in the United States, and Hol-

combe (1999) notes that the evolution of veterans benefits after the Civil War opened the political doors to transfer programs in the twentieth century.

If government expenditures are limited to certain items specified in the fiscal constitution, and if the beneficiaries of those expenditures pay the taxes to finance them, explicit redistribution might be prevented. However, when it becomes clear that rent-seeking activities can shift taxes away from the rent-seekers, toward others, and that rent-seeking activities can generate government production that benefit the rent-seekers at the expense of others, there is no principle to separate this type of redistribution in-kind from direct transfers. Lobbyists always wrap their requests in the cloth of the public interest. No rent-seeker says that he wants a tax cut to line his own pockets; rather, the argument is that the tax cut furthers the public good in some way. Similarly, no rent-seeker says that using public money to finance an activity will benefit the rent-seeker at the expense of others; rather, she says that the public production would enhance the public welfare. This allows the fiction that special interest benefits are intended to benefit the general public for a while, but when some rent-seekers benefit, more rent-seeking is encouraged, to the point where it becomes apparent that taxation and production policies are designed to produce benefits for some interest group, financed by costs imposed on the general public. Once this is recognized, it is a small step to simply make direct transfers. Once redistribution has become an explicit activity of the state, it attains legitimacy, to the point where academics (e.g., Musgrave 1959) refer to it as one of the state's primary functions.

One problem with the government's redistributional activities is that because government is not run by an omniscient and benevolent despot, but rather by democratic decision-making, there is no underlying principle that determines how redistribution will occur. Rather, as Stigler noted, benefits are transferred to those with political power. This leads interests to engage in rent-seeking, with the attendant political costs. And because benefits go to those with political power rather than to those who, according to some justification, deserve them, it is unlikely that the ultimate result of government redistributional activity will satisfy any reasonable criteria for efficiency or equity. Meanwhile, the rent-seeking and political costs remain.

The critical point to recognize is that regardless of the lofty goals or redistribution policy, the benefits go to those with political power, so unless one's objective is to enrich those with political power, using government for redistributive purposes will not accomplish one's ends. Policies cannot work, except toward that one end, but they impose substantial costs on everybody because of the excess burden and political costs associated with them. Samuelson (1956) attempted to justify the use of social indifference curves as an analytical device by arguing that government must be optimizing something, and whatever that is becomes the social welfare function. Samuelson is optimistic about the

efficiency of government production, along the lines of Becker (1983). Is it not possible that government acts in arbitrary and seemingly contradictory ways (for example, by engaging in public anti-smoking campaigns while subsidizing tobacco farmers)? However, to the extent that government as a single entity does actually maximize something, in a democracy the social welfare function becomes the interests of the majority. Redistribution becomes an activity to transfer benefits to those with political power. Political institutions prevent it from attaining more idealistic goals that some envision.

As with taxation and production, the solution, in theory, is to establish a fiscal constitution that prevents government from engaging in redistribution. This theoretical solution cannot, however, be practically applied. People who have government power can use that power to provide benefits to themselves financed by imposing costs on the citizenry at large, as Holcombe (1980) explains, eroding any fiscal constitution that is in place. Even if most people would not use government power that way, some would, and government positions will be more attractive to those who are willing to use them for personal gain. Thus, Brennan and Buchanan (1985) argue that government should be modeled as run by purely self-interested maximizers even if all people are not so motivated. If government power is there, some people will use it to further their personal ends at the expense of the general public. One way to limit the amount of rent-seeking is to limit the size and scope of government. A small government has less to offer rent-seekers than a large one, so rent-seeking activities and political costs associated with small government will be less than with large government. A rigid fiscal constitution is a good solution in theory, but a hard one to implement in practice.

12.   THE POLITICAL MARKETPLACE

The analysis in this chapter has offered a different way of viewing government's role in taxation, production, and redistribution, with policy implications that are considerably at odds with those of traditional public finance. In part, traditional public finance arrives at different conclusions because public sector economics has consistently ignored the political process through which public policy is produced. Once the analysis incorporates the political decision-making process, policy conclusions change dramatically. However, there is another line of reasoning that takes into account the political process, but arrives at the conclusion that democratic decision-making produces outcomes that approximate the efficiency of market outcomes. Following Becker (1983) and Wittman (1989), there is a literature that argues that the political marketplace allows individuals to register their demands in much the same way as the private market, and that politicians have an incentive to weigh the demands on all sides of an issue and clear the market by producing policies that accomplish collective goals at the lowest cost. If they did not produce public policies

at the lowest possible cost, they would be wasting some surplus that a political entrepreneur could use to offer a new policy that would make everybody better off. This line of reasoning directly challenges the conclusions of this chapter.

These models of efficient political exchange fail to take account of two factors that make political exchange significantly different from exchange in the market. First, political exchange requires the agreement of others in a way that has important differences with market exchange. Second, in politics, people can use the force of government to take resources away from others without their consent, whereas in market exchange, people give up resources only when they agree to do so because they are getting something they value more in exchange. When these two factors are taken into account, it is apparent that political exchange must be less efficient than market exchange.

13.   AGREEMENTS IN POLITICS AND IN MARKETS

In politics, nobody can undertake an action unless others also agree that the action is desirable. Under simple majority rule, a majority must agree before action can be taken. More complex democratic institutions allow more sophisticated political exchanges, but the principle remains that in politics, individuals require the cooperation of others to act. Even in a dictatorship, the dictator requires the support of others (the military, perhaps), so others must agree that the proposed action is desirable before it can be taken. In markets, nobody need agree on the desirability of an action, as long as the actor has the resources to finance it. Consider an example.

When Steve Jobs and Steve Wozniak designed their personal computer, they tried to sell the idea to existing computer companies, but those established companies thought that the product would not have a sufficient market. Because others did not agree with the merits of their idea, they started their own company, Apple Computer, to manufacture and market the computers themselves. With market exchange, nobody with whom they dealt had to think that Apple Computer was a viable company. Customers only had to want a computer more than they wanted the money Jobs and Wosniak were charging, and suppliers would gladly sell them parts regardless of whether they thought the business could succeed. Even if every customer thought, "I don't think this is a viable business, but if they're willing to sell me a computer for that price, I'll buy it," and if every supplier thought, "I don't think they can make money selling personal computers, but as long as they pay me I'll sell them parts," the market system would allow them to pursue their business. In the market, people can pursue activities without anybody else's support or agreement. In politics people need the support of others to proceed.

This is significant because the way politicians gain the support of others is by trading support on issues they care little about for support on issues that are

important to them. Because politicians have a limited number of votes they can trade, they must use them to their best advantage. That means trading their votes for support on issues that benefit them personally, rather than supporting the general public interest. Thus, as explained in more detail in an extensive literature,[6] programs tend to favor special interests over the general public interest. If a politician attempts to gain support for legislation in the general public interest, that means trading away votes that could be used on issues that would more narrowly benefit the politician. The fact that to accomplish something in politics, the support of others must be enlisted, means that politicians must conserve their political capital to support special interest issues. Thus, political exchange is less efficient than market exchange.

14.   COERCION

A second factor that creates inefficiency in the political marketplace is the fact that government action is based on coercion. In private markets, nobody can obtain a person's resources without that person's consent. In politics, government uses coercion to obtain resources and to enforce policies. As a result, there is a much higher information requirement in political markets than in private markets. In private markets, if people do not believe they are sufficiently informed to make a profitable trade in a market, they do not have to trade in that market. Some people invest in diamonds, others invest in art, and some people invest in real estate. People with little information about these markets can choose not to trade in them, and in the private market there is no way for a diamond dealer or an art dealer or a real estate broker to take resources from people without their consent. Nobody forces people to trade in markets they want to avoid. In markets, people only give up their resources if they agree to do so.

In politics the situation is different. If the agricultural lobby pushes for farm price supports, people who are uninformed about those price supports will end up paying the cost of financing them. In private markets, farmers could not take money from the general public without the general public tendering it voluntarily, but in political markets, farmers can use the force of government to take people's resources, not only without the consent of those who pay, but often without their knowledge. Uninformed people in a political markets are targets for having resources transferred away from them. This is why political markets are less efficient than private markets.

In private markets, where transactions are voluntary, people gain from production and exchange. In political markets, where some people can gain by using the power of government to forcibly take resources from others, people gain from predation. The key difference between private markets and political markets boils down to just this: production versus predation. Private markets

encourage people to be productive, which enhances overall wealth, whereas political markets encourage people to engage in predation: rent-seeking activity that transfers resources from some to others by force of government. Because of the coercive nature of government, it engages in activities that are different in nature from those of the private market, and therein lies the primary source of government inefficiency.

The hypothesis that political markets are as efficient as private markets, as promoted by Becker (1983) and Wittman (1989), implies that impediments to efficiency are limited to transactions costs and information costs, and then shows how political exchange can overcome these impediments, implying that political exchange is efficient. The inefficiency of political markets is not merely a matter of limiting agency costs, information costs, and so forth. Rather, its major source is the predatory power of government, which changes the nature of political activity when compared to market activity. In private markets, economists recognize that one prerequisite for efficiently performing markets is protection of property rights to encourage exchange rather than predation. In political markets, not only is this not recognized, political predation is elevated by some (Musgrave 1959) as one of the major functions of government. The answer to the literature which argues that political markets are as efficient as private markets is that predation is not as efficient as production.

The solution to this inefficiency is to design a system that limits predation and eliminates the pay-off to rent-seeking. The earlier analysis of taxation, production, and redistribution, was intended to sketch out how that might be done.

15.   TAXATION, PRODUCTION, AND REDISTRIBUTION

Taxation, production, and redistribution are core areas of inquiry in traditional public finance. In the theory of public finance as it developed in the last half of the twentieth century, the emphasis in all these areas has been on designing government policies to achieve a Pareto optimal allocation of resources, under the assumption that government has perfect information about preferences and technology (production functions). Armed with this information, economic analysis produces policies that would result in a Pareto optimal allocation of resources. There are many problems with undertaking public policy analysis in this manner. The analysis in this chapter has focused on one of them. Public policies are designed through a democratic political process, and are not imposed by a benevolent despot, and this chapter has shown that the optimal policies for taxation, production, and redistribution are significantly different once the analysis takes into account the political process that generates public policies.

There are additional problems with this traditional line of analysis. It assumes that government has information about preferences and technology that it could not obtain, so even a benevolent despot could not actually implement the policies it advocates. In some cases (Samuelson, 1954; Mirrlees, 1971) the analysts use social welfare functions that require interpersonal utility comparisons to find the welfare-maximizing policy, and assume that coercing people in order to try to achieve the optimal allocation of resources gives them no disutility, and creates no excess burden in the form of activities to try to avoid the coercion (Hochman and Rodgers, 1969). Even if the political problems analyzed in this chapter were solved, the policy implications from mainstream public finance would still be on shaky ground. This chapter has looked beyond these other problems, however, to focus on the political process within which public policies on taxation, public production, and redistribution are determined.

The fundamental difference between the analysis in this chapter and the traditional public finance analysis of taxation, production, and redistribution, lies in the nature of political exchange. Political exchange is different from market exchange in important ways, and while individual trades do take place in politics, it is collective decision-making and not individual exchange that results in public policies. Two legislators can agree to exchange votes on an issue, but that, by itself, does not create a public policy decision. A legislator may agree to support a lobbyist's proposal, but that, by itself, does not create a public policy decision. Rather, after all of these political exchanges have been made, public policy is created as a result of a collective decision. The nature of political exchange orients the political process toward providing benefits for special interests rather than furthering the general public interest.

The fact that interest groups can benefit from political decisions creates the incentive for rent-seeking activity. Rent-seeking can benefit the rent-seekers only because of the coercive nature of government policies. Rent-seekers attempt to use the coercive power of government to transfer resources from others to themselves. Despite the superficial similarity, political exchange is different from market exchange in fundamental ways that alter the way that resources are allocated. Rent-seeking activity creates political costs and misallocates resources so that even if government policy-makers could know the optimal allocation of resources, the political process still would not generate that optimal result. Once the political system that generates public policy is taken into account, optimal public policy changes in significant ways.

Optimal tax theory analyzes how tax systems can be designed in order to minimize the deadweight loss of taxation, and more recently, following Mirrlees (1971), how taxes should be apportioned among people to enhance social welfare. Optimal tax theory completely ignores political costs and assumes

that once economists show policy-makers the optimal tax policy, they can cost-lessly impose it. When the political system is taken into account, the analysis shows that rent-seeking imposes substantial costs on the economy, and that policies chosen through the democratic process will not be optimal. The well-known Ramsey rule for commodity taxation concludes that the optimal com-modity excise tax structure sets tax rates on goods in inverse proportion to their elasticities of demand. The analysis in this chapter concludes that once the political decision-making process is accounted for in the analysis, it is more likely that optimal commodity taxation implies the same tax rate for all goods rather than different rates for different goods. The problem is that when rates can vary among goods, that creates the incentive for rent-seeking among taxpayers to lower the rates they pay. Political costs can be minimized by de-signing a fiscal constitution that makes it difficult for tax rates to change, and that requires a substantial consensus for any changes to occur.

By the same logic, proportional income taxation that taxes all income is optimal once the political system is taken into account. This conclusion is substantially the same as that of the optimal income tax literature, even though the methodology is considerably different, and even though the optimal in-come tax literature completely ignores the political decision-making process. The fact that substantially different methodologies arrive at the same policy conclusion should make proportional taxation look even more desirable as a policy. This analysis showed that proportional income taxation is the only in-come tax structure that could be incorporated into a fiscal constitution based on a generally-acceptable principle, in order to minimize rent-seeking and po-litical costs. The general conclusion with all types of taxation is that an optimal tax structure is one that is embodied into a fiscal constitution in such a way that it is difficult to change, and can be changed only with a consensus agreement. This policy conclusion is substantially at odds with standard public finance, which ignores the political decision-making process that produces tax policy, and shows the importance of incorporating political institutions into the analy-sis.

When the political process is explicitly incorporated into the analysis of public production, similar conclusions emerge. Despite a well-developed the-ory on public policy to correct market failure through government production and regulation, once political institutions are incorporated into the analysis the conclusions change radically. As with optimal tax theory, political decision-makers cannot have sufficient information to make Pareto optimal production decisions as the theory describes, but even if the information were available, the incentive structure of democratic politics would prevent government from allocating resources optimally anyway. If government has enough discretion to allocate resources optimally as described in the public goods literature, then

its political institutions will be open to rent-seeking, which will lead to production for the benefit of special interests rather than in the general public interest. As with taxation, the optimal policy for public production is to sharply limit the discretionary abilities of government through a strict fiscal constitution that limits public production only to a small enumerated list of activities. In addition to limiting the inefficient use of resources in the production of special interest benefits, it also reduces rent-seeking and the attendant political costs.

The results of taking account of the political decision-making process in redistribution leads to similar conclusions. The nature of the political process will prevent government from accomplishing its redistributional goals anyway, unless those goals are to transfer resources from the general public toward those with political power. Yet the possibility of obtaining transfers creates rent-seeking opportunities and the political costs associated with them. The optimal policy toward government redistribution is to create a constitutional prohibition against it. As Higgs (1987) notes, the American Founders tried to do that, and until the end of the nineteenth century American government operated under a fiscal constitution that prohibited redistribution. As taxation and public production became increasingly distributional in nature, however, the door was opened to explicit transfers, so that by the middle of the twentieth century redistribution was accepted as one of the core functions of government.

Recognizing the problem, what can be done to create an effective fiscal constitution to limit rent-seeking and political costs, and to redirect public sector activity from producing benefits for narrow special interests toward the general public interest? There is no easy answer to the question; if there was, the solution would already have been implemented. However, a part of the problem is that much public policy analysis takes place without explicitly recognizing the problems inherent in democratic decision-making. Thus, many analysts do not even see that these problems exist. Some progress can be made simply by promoting the type of public choice analysis contained in this chapter. Is that wishful thinking? In the conclusion of his *General Theory,* John Maynard Keynes (1936, p. 383) argued, "… the ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed, the world is ruled by little else." The provides the justification for undertaking this type of analysis, and for taking it seriously.

NOTES

1. This statement would not have been true a few decades earlier, as even well into the 20th century, reputable economists argued the merits of central planning. For example, Samuelson (1973, p. 883) said that even though the U.S. had roughly double the per capita income of

the Soviet Union at that time, the Soviet Union's superior centrally planned economy made it more productive, so that Samuelson projected that perhaps as soon as 1990, and almost surely by 2010, the Soviet Union would have caught up with the U.S.

2. Not all economists believe that government is necessary or desirable. See, for example, Rothbard (1973), which, interestingly enough, has the same publication date as Samuelson's book cited in the first footnote.

3. But again, see Rothbard (1973) and Benson (1990) for arguments that these functions could be undertaken by the market if the government did not take them over by force.

4. See, for examples, Tiebout (1956), Minasian (1964), and Holcombe (1997a, b) for critiques of Samuelson's public goods theory.

5. Although, as Minasian (1964) notes, market production may generate such information in cases where government production would not. Hayek (1945) makes a very persuasive case that the market produces and reveals information to economic agents in a way that could not be accomplished in the absence of a market.

6. See, for examples, Weingast, Shepsle and Johnsen (1981) and Holcombe (1985).

REFERENCES

Akerlof, George A. (1978). "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Training." *American Economic Review* 68 (1): 8-19.

Anderson, Terry L. and Hill, Peter J. (1980). *The Birth of a Transfer Society.* Stanford: Hoover Institution Press.

Bator, Francis M. (1958). "The Anatomy of Market Failure." *Quarterly Journal of Economics* 72: 351-379.

Becker, Gary. (1983). "A Theory of Competition Among Pressure Groups for Political Influence." *Quarterly Journal of Economics* 98 (3): 371-400.

Benson, Bruce L. (1990). *The Enterprise of Law: Justice Without the State.* San Francisco: Pacific Research Institute.

Blackorby, Charles and Donaldson, David. (1988). "Cash Versus Kind, Self-Selection, and Efficient Transfers." *American Economic Review* 78 (4): 691-700.

Brennan, Geoffrey and Buchanan, James M. (1985). *The Reason of Rules: Constitutional Political Economy.* Cambridge: Cambridge University Press.

Brennan, Geoffrey and Lomasky, Loren. (1993). *Democracy and Decision: The Pure Theory of Electoral Preference.* Cambridge: Cambridge University Press.

Bruce, Neil and Waldman, Michael. (1991). "Transfers in Kind: Why They Can Be Efficient and Nonpaternalistic." *American Economic Review* 81 (5): 1345-1351.

Buchanan, James M. (1967), *Public Finance in Democratic Process.* Chapel Hill: University of North Carolina Press.

Buchanan, James M. (1975). "Public Finance and Public Choice." *National Tax Journal* 28 (4): 383-394.

Buchanan, James M. (1976). "Taxation in Fiscal Exchange." *Journal of Public Economics* 6 (1/2): 17-29.

Buchanan, James M. and Tullock, Gordon. (1962). *The Calculus of Consent.* Ann Arbor: University of Michigan Press.

Friedman, Milton. (1962). *Capitalism and Freedom.* Chicago: University of Chicago Press.

Hayek, Friedrich A. (1945). "The Use of Knowledge in Society." *American Economic Review* 35 (4): 519-530.

Hayek, Friedrich A. (1960). *The Constitution of Liberty.* Chicago: University of Chicago Press.

Hettich, Walter and Winer, Stanley L. (1988). "Economic and Political Foundations of Tax Structure." *American Economic Review* 78 (4): 701-712.

Higgs, Robert. (1987). *Crisis and Leviathan: Critical Episodes in the Growth of American Government.* New York: Oxford University Press.

Hochman, Harold M. and Rodgers, James D. (1969)."Pareto Optimal Redistribution." *American Economic Review* 59: 542-557.

Holcombe, Randall G. (1980)."A Contractarian Model of the Decline in Classical Liberalism." *Public Choice* 35 (3): 277-286.

Holcombe, Randall G. (1985). *An Economic Analysis of Democracy.* Carbondale: Southern Illinois University Press.

Holcombe, Randall G. (1992). "The Distributive Model of Government: Evidence from the Confederate Constitution." *Southern Economic Journal* 58 (3): 762-769.

Holcombe, Randall G. (1994). *The Economic Foundations of Government.* New York: New York University Press.

Holcombe, Randall G. (1997a). "Selective Excise Taxation from an Interest Group Perspective." In *Taxing Choice: The Predatory Politics of Fiscal Discrimination,* Shughart, William F. II, ed. New Brunswick: Transaction Publishers.

Holcombe, Randall G. (1997b). "A Theory of the Theory of Public Goods." *Review of Austrian Economics* 10(1): 1-22.

Holcombe, Randall G. (1998). "Tax Policy from a Public Choice Perspective." *National Tax Journal* 51 (2): 359-371.

Holcombe, Randall G. (1999). "Veterans Interests and the Transition to Government Growth: 1870–1915." *Public Choice* 99 (3–4): 311-326.

Kreuger, Anne O. (1974). "The Political Economy of the Rent-Seeking Society." *American Economic Review* 64 (3): 291-303.

Lindahl, Erik. (1967). [orig. 1919]. "Just Taxation—A Positive Solution." In *Classics in the Theory of Public Finance,* Musgrave, Richard A. and Peacock, Alan T., eds. New York: St. Martin's.

Minasian, Jora. (1964). "Television Pricing and the Theory of Public Goods." *Journal of Law & Economics* 7: 71-80.

Mirrlees, James A. (1971)."An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38: 175-208.

Mirrlees, James A. (1976). "Optimal Tax Theory—A Synthesis." *Journal of Public Economics* 6 (4): 327-358.

Musgrave, Richard A. (1959). *The Theory of Public Finance.* New York: McGraw-Hill.

Ramsey, Frank P. (1927). "A Contribution to the Theory of Taxation." *Economic Journal* 37: 47-61.

Rothbard, Murray N. (1973). *For a New Liberty.* New York: Macmillan.

Samuelson, Paul A. (1954). "The Pure Theory of Public Expenditure." *Review of Economics and Statistics* 36: 387-389.

Samuelson, Paul A. (1955). "A Diagrammatic Exposition of a Theory of Public Expenditure." *Review of Economics and Statistics* 37: 350-356.

Samuelson, Paul A. (1956). "Social Indifference Curves." *Quarterly Journal of Economics* 70: 1-22.

Samuelson, Paul A. (1973). *Economics,* 9th ed. New York: McGraw-Hill.

Slemrod, Joel. (1990). "Optimal Taxation and Optimal Tax Systems." *Journal of Economic Perspectives* 4 (1): 157-178.

Slemrod, Joel and Sorum, Nikki. (1984). "The Compliance Cost of the U.S. Individual Income Tax System." *National Tax Journal* 37 (4): 461-474.

Smith, Adam. (1937) [orig. 1776]. *The Wealth of Nations.* New York: Modern Library.

Stigler, George J. (1970). "Director's Law of Public Income Redistribution." *Journal of Law & Economics* 13(1): 1-10.

Tiebout, Charles M. (1956). "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64: 416-424.

Tullock, Gordon. (1967). "The Welfare Cost of Tariffs, Monopolies, and Theft." *Western Economic Journal* 5 (3): 224-232.

Weingast, Barry, Shepsle, Kenneth A., and Johnsen, Christopher. (1981). The Political Economy of Benefits and Costs: A Neoclassical Approach to Distributive Politics." *Journal of Political Economy* 89 (4): 642-664.

Wicksell, Knut. (1967) [orig. 1896]. "A New Principle of Just Taxation." In *Classics in the Theory of Public Finance,* Musgrave, Richard A. and Peacock, Alan T., eds. New York: St. Martin's.

Winer, Stanley L. and Hettich, Walter. (1998). "What is Missed if We Leave Out Collective Choice in the Analysis of Taxation?" *National Tax Journal* 51 (2): 373-389.

Wittman, Donald. (1989). "Why Democracies Produce Efficient Results." *Journal of Political Economy* 97 (6): 1395-1424.

Chapter 7

PUBLIC REVENUE FROM LAND RENT

Fred E. Foldvary
Santa Clara University
ffoldvary@scu.edu

Abstract   Rent, the return on land, can be taxed without creating an excess bur-
den. The geographical benefits of a location, especially its civic services
and public works, become capitalized into higher land value, while the
taxation of the rent is capitalized into lower prices for purchasing land.
The taxation of site rentals thus returns to the government the rentals
generated by its services. In principle, the optimal expenditure for a pub-
lic good impacting territory is where the induced marginal rent equals
the marginal cost of the good. This chapter provides an analysis of the
ethics, economics, and politics of basing public revenue on land rent, its
consequences for an economy and the public well being, and a public-
choice analysis of why governments have not widely adopted this type
of revenue.

Keywords: Excess burden, land rent, public revenue, public choice analysis

JEL classification: H20, H71, R51

Rent, the return on land, can be taxed without creating an excess burden. The
geographical benefits of a location, especially its civic services and public
works, become capitalized into higher land value, while the taxation of the
rent is capitalized into lower prices for purchasing land. The taxation of site
rentals thus returns to the government the rentals generated by its services.
In principle, the optimal expenditure for a public good impacting territory is
where the induced marginal rent equals the marginal cost of the good. Using
rent to finance the fixed costs of declining-cost services such as transportation
complements marginal-cost pricing. Using rent for public revenue instead of
taxing productive activity also has beneficial macroeconomic effects on urban
land use, economic growth, and even business cycles. The concept that land
rent can serve as an efficient and equitable source of public revenue has been

recognized by economists including the Physiocrats, Adam Smith, David Ricardo, Henry George, Harry Gunnison Brown, and contemporary economists, and has been practiced in a number of localities. Private communities are also in effect using land rent for their public revenue.

Adam Smith (1776b [1976], p. 370) stated that "Ground rents, and the ordinary rent of land, are, therefore, the species of revenue which can best bear to have a particular tax imposed upon them." Many economists have agreed with this policy proposition, but governments have not adhered to it, instead mostly taxing productive activity. Although rent-based public revenue has major implications beyond financing government, affecting topics from business cycles to poverty to urban sprawl, mainstream macroeconomics and even public economics has yet to integrate Smith's proposition into its theory and policy prescriptions. Some economists have been critical of or have disagreed with the proposition or the implications as discerned by advocates such as Henry George (Andelson, 1979a, b). This chapter provides an analysis of the ethics, economics, and politics of basing public revenue on land rent, its consequences for an economy and the public well being, and a public-choice analysis of why governments have not widely adopted this "species of revenue."

1.   THE ECONOMICS OF RENT-BASED PUBLIC FINANCE

1.1.   Land as a Factor of Production

A "factor" is a category of resource inputs. The three classical factors of production are land, labor, and capital goods. "Land" includes all natural resources and natural opportunities, nature being all resources prior to and apart from alteration by human action. Besides the space surrounding the earth, land includes material natural resources (minerals, oil, water), the electro-magnetic spectrum, and wildlife.

Real-estate land includes the space at, above, and below the surface of the earth, as well as the attached natural material features and resources such as water, plants, and soil. While the materials may be altered by human action, after which they are no longer economic land, the space itself cannot be moved, used up, altered, or expanded. Sites are fixed in location and in extent, and so the supply of this land is fixed, inelastic in supply. The quantity of space does not rise or fall with changes in its price.

Non-renewable material resources such as minerals, coal, and oil are subject to depletion, and exploration can discover previously unknown reserves. Wildlife and water are prime examples of land that can be expanded or contracted, depending on the rate of depletion and the preservation or enhance-

ment of the habitat or sources. Managed and controlled living beings or water are, however, no longer land, but capital goods.

Capital goods are produced or altered goods that are in turn used to produce more wealth rather than directly consumed by the final user. Besides tools and buildings, capital goods include civic infrastructure such as streets, parks, and utilities (pipes, wires, lighting). The alteration of a site to facilitate construction is also a capital good, so that in leveling or draining a site, such improvements are capital goods rather than being part of the natural land. Such improvements typically require continuous maintenance involving labor. The revenue generated by a site includes the natural land rent plus the extra rental induced by the civic goods plus wages for the labor associated with the improvements. Tideman (1994b) distinguishes three returns, (1) the value attributable to nature; (2) the value attributable to public services, and (3) the value attributable to private activities. The distinction between the original land rent and payments for capital goods expended on the land was recognized by John Stuart Mill (Lackman, 1977), although like Henry George later, Mill thought that such capital goods that permanently become part of the site and do not require maintenance lose the character of capital goods and become, in effect, part of the land, its return being rent (Mill, 1908, p. 408).

The term "rent" strictly is the ground rent, the return due to the land as the natural features, while "rental" refers more generally to the return on the capital goods attached to land and to the wages included in such rental payments, and also to a payment that includes both rental and rent. "Land" strictly refers to the natural resource factor, and "site" refers to a plot of land plus improvements attached to the land. But more loosely, "rent" is used for "rental" when the context is clear, and similarly "land" for the site.

In classical economic analysis, especially by David Ricardo and Henry George, land is divided into various grades reflecting its natural productivity such as due to location or fertility. Since labor and capital good are mobile, if they are presumed to be uniform, they have a uniform return. The least productive land in use is called the "margin of production." Submarginal land is not used and therefore has no rent. The wage level at the margin determines wages for the rest of the economy, and after paying for capital goods, the surplus is land rent. If the margin moves out to less productive land, the productivity at the margin falls, hence the whole wage level falls, and the rent of the supermarginal lands rise. Thus is the Ricardian differential rent determined.

The modern neoclassical notion of the marginal product of land, the extra production when a bit more of land of the same quality is used, keeping the other factors constant, is consistent with the concept of differential rent. Classical theory provides the added insight that the market rental is not just determined by how productive a site is, but how productive it is relative to the relevant less productive sites. Nicolaus Tideman (1999) adds the insight that

the rent of land "should be defined as the opportunity cost of leaving unimproved land unused."

Some neoclassical economists, such as John Bates Clark, have argued that land has the same marginal properties of the other factors and that land does not have properties that uniquely distinguish it from capital goods, stating that the returns to any factor could be viewed as a differential return or surplus (Collier, 1979).

However, Harry Gunnison Brown (1917) pointed out that while the interest return on capital is computed as the interest rate times the price, the return on land is "an absolute amount measured and determined by the surplus over production on the extensive or intensive margin. It is not determined by the value of land" (1980 [1917], p. 3). Rather, the price of land is the discounted value of future rents. Moreover, unlike labor and capital goods, land has no opportunity cost of production, there being no leisure alternative nor any opportunity cost of resources used. Gaffney (1994a) further points out several distinctions between spatial land and capital goods, land being nonreproduceable, permanent and recyclable, fixed in supply, immobile in space and uncontrollable in time, does not turn over like capital good do, and is not directly convertible into capital goods (a consequence of its not turning over).

1.2.   The Capitalization of Externalities, Costs and Benefits

An "external effect" or "externality" is an uncompensated cost or benefit. Civic works and services such as streets, parks, security, fire protection, public transit, and schools increase the demand to be located in the affected territory. This increases the site rentals and thus also the site prices. If the owners of the sites do not pay for these civic goods, then there exists a positive external effect that benefits them, at the expense of the taxpayers. Civic goods funded from tax sources other than site rentals therefore generally result in a forced transfer of income and wealth from taxed workers and owners of capital goods to site owners.

The externalities of land were recognized by Adam Smith, who wrote on how "every improvement in the circumstances of society tends either directly or indirectly to raise the real rent of land" (Smith, 1776a [1976], p. 275). Henry George recognized the "tragedy of the commons" long before Hardin's (1968) 20th century phrase become popular, and that his remedy, a tax on the rent, would eliminate the congestion externalities (George, 1879, pp. 328, 397-406; Dwyer, 1982, p. 298). George also recognized that externalities are ubiquitous: "No one can keep to himself the good he may do, any more than he can keep the bad. Every productive enterprise, besides its return to those who undertake it, yields collateral advantages to others … And in the value or rent of

land is this general gain expressed in a definite and concrete form" (George, 1879, pp. 435-436). Since site rentals reflect the net territorial externalities, "establishing private tenure but at the same time collecting the competitively determined economic rent for the public benefit will establish an optimal level of externality" (Dwyer, 1982, p. 299).

"Capitalization" is the creation (or reduction) of asset value by a stream of income (or expense). The market value of such an asset is the net present value of the stream. Since civic public goods generate a stream of value to site owners, civic services become capitalized into site rentals. This capitalization is due to the fixity of land. For produced goods and movable factors, an increase in profit induces a greater supply of the goods either through greater production or importation. But since space cannot be expanded or imported, the result of the increased profit is only a higher rental.

The simplified formula for the relationship between asset value and the stream of income is the net return divided by the sum of the interest and tax rates. For sites, it is

$$p = r/(i + t),$$

where p is the price of a site, i is the real interest rate (excluding the inflation component), and t is the tax rate on p. For example, if the rental is \$10,000 per year, the interest rate is 5%, and the tax rate based on the price is 15%, then the price would be

$$\$10000/(0.05 + 0.15) = \$10000/0.2 = \$50,000.$$

This assumes that $r, i,$ and t are not anticipated to change. The rental must both pay for the tax (t times p) and the normal return on an asset (i times p).

If r is the pure land rent, excluding payments for the site improvements, then a tax on the rent does not affect the maximal market rent r, but only the price p. If the owner was already charging the tenant what the market could bear, then the tax cannot be passed onto tenants, since a higher rent will just move them to use a lower amount of space, creating vacancies. Landlords will then decrease the rent to that amount that clears the excess supply. Since land has no cost of production, its price decreases when taxed, unlike goods in current production that have costs of production that cannot be decreased.

Therefore the current owner bears the entire burden of a tax on rent, by not keeping all the rent. The tenant's rent pays the tax but is not a burden to the tenant, since he pays the same rent whether it is taxed or not. After the tax or charge is in effect, a new landowner does not really have a burden, because the tax will be offset by the lower price of land. Hence it is only the owner at the time the tax rate is increased that has a burden. Once in effect, a tax on rent

has no burden at all, neither on society nor the individual title holders Richman (1973).

There is a huge excess burden from taxing labor and capital, the burden in the U.S. estimated at more than $1 trillion for 1998 Tideman and Plassmann (1998). Henry George's remedy for this waste of resources is to shift taxation, eliminating taxes on all productive activity, instead basing general public revenue on site rentals. This shift would substantially raise wages, both because wages would not be taxed and because the wage level as a whole would rise with the increased productivity of the economy.

The Georgist tax policy is the ultimate in supply-side economics and policy: it reduces the marginal tax rates to zero. The tax on land rent has the same effect as a lump-sum tax: there is no tax penalty on any additional labor, production, and investment, nor any arbitrary fiscal discrimination. With regulation being a type of tax, or a substitute for taxation, the "single tax" on site rentals, eliminating all other imposed arbitrary costs, would also eliminate excessive regulations, maximizing the supply-side production of goods and growth-maximizing investment. With "civil" asset forfeiture also being in effect a tax when the person affected is not convicted (Boudreaux and Pritchard, 1997), a pure single tax would also rule out such confiscations. It would, however, not exclude fines, penalties, and other compensation for damages. This single tax would thus only be a tax in form, since in substance, compensation for damages and for the use of public resources are user payments.

Georgist policy also strengthens rather than impinges on rights of possession. In its purity, the Georgist program splits the ownership bundle of rights to land into two basic strands, rights of possession and rights to the return. Only the return consisting of economic rent is shared or collected by the community; the title holder otherwise has complete rights of possession in his domain (see Pullen, 2001, and other comments in the same issue). That domain would normally exclude high-up airspace for airplane routes and possibly public-access easements.

1.3.   Speculation, the Margin of Production, and Land Use

Henry George (1879) added expectations to the Ricardian classical rent theory. He theorized that when growth is anticipated, speculators will obtain land expecting the rent and price to rise, and the present-day price will rise both in anticipation of the future price and due to the speculative demand added to the demand for use. Real estate can then become priced too high to make current investment profitable, which reduces the demand for investment. Reduced demand then spreads throughout the economy, leading to a recession and depression. Moreover, development may shift from urban centers to the

fringes as the center is priced too high to be used profitably and developers at the fringe anticipate gains from advances in those site values. This phenomena is intensified when government finances public works from taxes on wages, profits, and capital goods, since the gain in site value is boosted by the future civic improvements which the landowners do not have to pay for.

It can be sound economics to await construction until a future time, so that the past construction does not have to be torn down. But such market-enhancing speculation can become market hampering when the financing of civic goods becomes a forced transfer from consumers, workers, and entrepreneurs to landowners. Often, landowners are not passive recipients of increasing land value, but are active rent seekers, influencing rental-enhancing legislation such as zoning and public works Foldvary (1998a).

Mason Gaffney (1994a, p. 93) recognizes two effects of market-hampering land speculation. In type A, buyers "force the future" by developing for future demands which do not materialize. In type B, landowners just hold land underused and "free ride on the future." The effective remedy for both types is site-value revenue collection.

Site-value taxation "is neutral with regard to socially optimal speculation" (Feder, 1996, p. 44), whereas taxes on labor and capital that capitalize up site values are distortive. B. M. Anderson in 1910 pointed out that if the tax is on economic rent regardless of its development, it does not alter the timing of construction. Whatever action maximizes the present value of returns without the tax continues to maximize the present value after a tax that takes fixed amounts at each time interval. Moreover, when markets are imperfect (e.g., when there are unarbitraged disparate beliefs about the optimal future use of land), a tax on rent is generally better than neutral, such as helping to prevent the "winner's curse" of auction markets, where the most optimistic bidder wins a bid Tideman (1999).

The public collection of the ground rent therefore can even have beneficial effects beyond the revenue obtained, providing an excess benefit. The market rental—what a site would fetch in auction rather than what a particular tenant may be paying—reflects the current externalities of the location and social cost of excluding others from that fixed site. When the title holder pays a rental reflecting the benefits of the land and of civic services, it induces a productive use of the site in order to pay the charge as well as generate a profit. This results in an efficient use of land. This can include postponing construction, but taking into account the explicit carrying cost. When the titleholder instead benefits from rising rentals from services he does not pay for, then his less intense use pushes residential and commercial uses away from the city center not because of sound economics but because of rent seeking.

1.4.   The Marginal Analysis of Public Spending and Site Rentals

The marginal cost of making something is the cost of one more unit of the good. The marginal rental generated by a territorial public good is the addition to the rentals of sites due to having a bit more of the good. We can think of the public good, such as a park, as providing a flow of services over time. If the usage is not congested, the use of the public good by a user does not detract from the use by the others; all have access to or use the entire good. We can distinguish between the availability of a public good, its potential or stand-by use, and its direct use (Holterman 1972). For many civic goods, such as security, the public aspect consists of its availability.

The public goods literature has asserted that when public goods are provided, people will try to be free riders, using the good without paying for it, because once it is available, there is no way to make them voluntarily pay what it is worth to them. But that overlooks the spatial dimension of public goods. Most goods and services typically provided by government are territorial; the usage is mostly by residents living and working in the territory. They have to pay rentals in order to be located there. The marginal rental generated by additional amounts of a wanted public good reflects the demand and the value placed on the good by the residents and users.

The users therefore are not free riders, since they pay for the service in rentals. If the amount of a collective good G can vary, then if an increase in G yields a greater increase in rental than the cost of the extra G, it is beneficial to provide ever more G until the marginal rental generated just equals the marginal cost of providing more. The optimal quantity of the public good is therefore the amount for which the marginal rental equals the marginal cost, if at that quantity the total rental generated by the good is greater than the total cost. That quantity maximizes the site rentals, similar to firms maximizing profits at the quantity where marginal revenue equals marginal cost.

Several economists have shown that under certain conditions, the site rentals equal the optimal cost of the public goods. Joseph Stiglitz (1977, 1983) has a model of a community that chooses the level of public goods that maximizes its site rentals. With site rentals paying for the public goods, site-value maximization leads to an efficient level of public goods.

Suppose there is a utility function $U(G, X)$, where G is a collective good, X is a severable good (a private good, consumed individually), and U represents the level of utility preference for that good relative to other goods. In the simple model of Atkinson and Stiglitz (1980), Y is output, a function of the number of workers N; capital goods are folded into labor. Therefore,

$$Y = f(N) = XN + \text{G},$$

and
$$X = (f(N) - G)/N.$$

The first-order condition, taking the first derivative to get the rates of change, implies

$$G = Y - N\partial Y/\partial N,$$

hence the optimal amount of the public good G equals total output (Y) minus total wage payments, since the wage equals the marginal product of labor, the change in Y for an extra worker, ∂N. Total output minus wages equals the site rentals. Hence, the population that maximizes consumption per capita is that for which the total rentals equal the expenditures for public goods. This result has been named the "Henry George theorem" (HGT) by Stiglitz (1977), since the site rentals constitute the "single tax" advocated by Henry George, which exactly and optimally finances G. The real world is more messy than that, but the HGT does logically back up Georgist theory.

William Vickrey had a similar conclusion. In a model of a city, Vickrey (1994 [1977], p. 342) finds that rents equal "the marginal social cost of land occupancy," which is based on "the increase in transportation costs involved in the occupancy of more land," and "total transportation costs will be equal to total land rents." Vickrey notes that this confirms the proposition by Harold Hotelling that the taxation of land rent appropriately finances the intramarginal cost residues of increasing-return industries pricing their output at marginal cost. In Vickrey's "GHV theorem" (George-Hotelling-Vickrey), land rents are just sufficient to pay the costs beyond the marginal costs paid by the users. Indeed Vickrey finds that it is necessary for efficiency for rent to finance such decreasing cost services, and if any of these rents are instead appropriated by private landowners, this reduces efficiency (p. 345).

Nicolaus Tideman (1985) notes that such conclusions involve some assumptions such as the mobility of labor and capital, public goods affecting a limited area, and the existence of persons outside the community who value the goods. But even if these conditions are not perfectly present, there is also land rent that would exist even without the public activities that can be tapped along with the generated rentals.

Critics of site-value taxation argue that the total generated rental would not be sufficient to finance all of government. Henry George (1879, p. 406) believed that the total rental would be sufficient, and in more developed economies, more than sufficient to finance public goods. It is quite possible for the total rental generated by a public good to be greater than its total cost at the optimal quantity. Studies of the impact of subways in New York City

show that landowners could have paid for it and still made a profit. As another example, the George Washington Bridge across the Hudson River increased New Jersey site values by six times the cost of the bridge (Tucker, 1958, p. 11).

The totality of land rent and site rentals is much greater than generally recognized. Mason Gaffney (1970) has estimated that site values are more than half of all real-estate market value. The Bureau of Economic Analysis estimated the total value of real estate in 1986 to be $10.8 trillion Miles (1990). That would put the site value of the US at over $5 trillion in 1986, and that excludes the rent of material land (oil, coal, minerals, water), the electro-magnetic spectrum, airline routes and satellite orbits, and the fact that much of present-day site value has been negatively capitalized due to taxes, regulations, excessive litigation, and other imposed costs that reduce profits.

One could argue that government should not be spending so much of public funds for services which either are not paid from user fees or which do not generate more rental than the cost. Much of government spending is for transfer payments rather than providing public goods, and many of these transfers are the result of political pressure groups seeking subsidies that do not benefit the public. Still, if additional funds other than user fees and site rentals are desired, other sources of funding can include profits from government-operated enterprises as well as charges on negative externalities such as congestion and pollution, and there is therefore no economic need to tax beneficial activities such as entrepreneurship, labor, investment, and the sale of goods.

1.5.   Free Trade and Land-value Taxation

Henry George's work *Protection or Free Trade* (1886) is a classic argument for free trade and against trade limitation. He took the case for free trade to its logical conclusion: "The mere abolition of protection … is such a lame and timorous application of the free-trade principle that it is a misnomer to speak of it as free trade … It applies as well to domestic as to foreign trade, and in its true sense requires the abolition of all internal taxes that fall on buying, selling, transporting or exchanging" (p. 286).

George advocated what he called "true free trade"—trade with no barriers against both domestic as well as foreign trade. Regarding the French Physiocrats, George wrote (1894, pp. 152-153) "In their practical proposition, the single tax, they proposed the only means by which the free trade principle can ever be carried out to its logical conclusion—the freedom not merely of trade, but of all other forms and modes of production, with full freedom of access to the natural element which is essential to all production."

1.6.   Land Value Taxation in Practice

The implementation of public revenue from site values requires an assessment of its market rent based on its best use. Assessors have several techniques for estimating this. First, data can often be obtained from sales or leases of unimproved land or sites with separate titles for land and buildings. Second, the value (or income) of a building, adjusted for depreciation, can be subtracted from the total property value (or income), the remainder being site rental. Third, assessors have maps, now computerized, with which they can extrapolate neighborhood values, adjusted for special features. Mason Gaffney found that some 80 percent of assessors' time was spent in assessing building values and only 20 percent in land values (Cord, 1979, p. 6). There are other techniques which can be included, such as self-assessment with buyouts when there is an offer above a certain premium.

Land has been a source of public finance for millennia world-wide; in primitive economies, it is a visible and ubiquitous resource. In the American colonies, the ownership of land was a key basis for a tax based on the faculty or ability of a resident to pay. Land grants were also a form of payment by the colonies and later, by the U.S. State and federal governments. During the 1800s, the States and local governments came to also tax improvements (Sakolski, 1957, p. 253). However, many countries and localities have taxed land values aside from improvements, including cities in Australia, New Zealand, South Africa, and Pennsylvania (Cord, 1979; Chandler, 1982; Andelson, 2000). In Pennsylvania and elsewhere, cities may split their property tax into separate rates for land and improvements, and where they have done so, those cities having lower rates on buildings have had greater growth Saunders (1999). In Japan during the latter 1800s and Taiwan after 1950, the national taxation of land rent was a significant element of their rapid economic development (Chandler, 1982; Harrison, 1983). Some places, such as Hong Kong, have obtained substantial revenue from site leaseholds. (For a survey on the implementation of land-value taxation world-wide, see Andelson, 2000.)

The most complete implementation of site value taxation occurred in Kiaochow (now Jiaoxian), the German colony in China from 1898–1915, whose main city was Tsingtao, now Qingdao. The imperial commissioner, Ludwig Wilhelm Schrameier, a member of the German Land Reformers, established a land-value tax of six percent (Silagi, 1984; Peterson and Hsiao, 2000). Sun Yat-sen, China's revolutionary leader, was impressed with this example, and proposed land reforms and land-value taxation, which would later be put into the Chinese constitution in 1930, but only implemented in some cities until the Nationalists came to Taiwan and implemented it nation-wide in 1950 (Chandler, 1982).

Besides Sun Yat-sen, other government leaders who favored public revenue from land rent included Aleksandr Kerenski (Russian prime minister of the

first revolution in 1917), Winston Churchill and UK prime minister Lloyd George, and Theodore Roosevelt (Chandler, 1982). Thus, during the early 20th century in the aftermath of the Georgist single-tax movement, leading officials of the major powers—the U.S.A., U.K., China, Russia—understood and favored site-value taxation, but for various reasons were not able to enact it. The movement lost its momentum during the turmoil of World War I, and never recovered its mass appeal.

2.   A BRIEF HISTORY OF RENT THEORISTS

2.1.   Land Theorists prior to Henry George

The taxation of land rent received substantial attention in early economic analysis. Sir William Petty recognized that a tax on rent would be capitalized in its price, and was favorably inclined to taxing land rent. In France, Pierre le Pesant, Sieur de Boisguilbert, proposed a single tax on land. The Dutch philosopher Spinoza also had that insight: "The fiscal base should be a single tax [on] the whole soil" (Chandler, 1982, p. 53).

This idea was carried forward by the French economic school of 1750–1790 that called itself "Physiocracy," meaning the rule of natural law, since they held that natural laws applied to society and an economy just as it does to the physical world. The physiocrats theorized that land creates a "net product" beyond payments for labor and capital goods. This net product, which we now recognize as rent, can be taxed without hampering production; therefore they proposed free trade and an *impôt unique* or single tax on the net product (Velde, 1997).

Adam Smith visited the Physiocrats in France, and he was influenced by them. Recognizing the principle of the capitalization of territorial externalities into site rents, he wrote that "every improvement in the circumstances of the society tends either directly or indirectly to raise the real rent of land, to increase the real wealth of the landlord" (1776a, p. 275). He also recognized rent as a surplus that is not due to the exertion of the title holder: "As soon as the land of any country has become private property, the landlords, like all other men, love to reap where they never sowed, and demand a rent even for its natural product" (1776a, p. 56). Smith also recognized that taxes on rent do not get passed on to tenants: "A tax on ground rents would not raise the rents of houses. It would fall altogether upon the owner of the ground rent" (1776b, p. 370).

Other public-rent proponents of the 1700s included William Ogilvie, professor of King's College in Old Aberdeen, Thomas Spence (1775), and Thomas Paine (1797) in his work, *Agrarian Justice* (Davidson, 1899; Schwartzman, 1997).

David Ricardo (1821) derived the "law of rent" as the residual of the fixed factor, a surplus remaining after paying for the mobile factors of labor and capital goods. Ricardo deduced that a tax on rent does not affect the prices of goods; the rent is high because the price of corn is high, rather than the price being high because the rent is high.

Patrick Edward Dove in *The Theory of Human Progression* (1850) and *The Elements of Social Science* (1854) foreshadowed Henry George on the "land question." In the latter work, Dove stated that there is a surplus beyond the costs of production, which is land rent. Workers "pay both rent *and* taxation, and consequently are *robbed,* for *robbery* it is" (Davidson, 1899, p. 65). Advocating a single tax on land rent, Dove stated: "National Property there must be *somewhere,* and assuredly it is more just to take that property from the natural value of the soil than from the individual fruits of labour" (p. 66). "It would make one simple tax" (p. 67).

John Stuart Mill, like Smith, recognized that the rental of land increases with the progress of society, and argued that the owners have no just claim on this increase. Mill regarded rent as a result of a monopoly; it is not the absolute monopoly of a single firm in an industry, but rather a monopoly in the classical sense of it being impossible to enter the field to expand the supply. A new owner must obtain land from a previous owner, being unable to produce new land. Altogether the landowners have a monopoly of the supply.

Herbert Spencer, the British philosopher, proposed in *Social Statics* (1850) the "law of equal freedom." From this basic concept he derived several principles, including that all persons have equal rights to the use of the earth, which he proposed to implement by leaseholds that paid rent to the government, with no taxes. In his later years, however, he retreated from this land stance, though not completely rejecting it.

2.2.   Henry George

The economist most identified with taxing land rent is Henry George. The central thrust of George's thought "is the insight that natural opportunity should be open on the same terms to all, and socially created values socially appropriated, while the fruits of private effort should be left inviolate to their producers" Andelson (b). He combined normative and positive theory, adding a moral dimension to the economic analysis, indeed creating a paradigm of thought that has been named Georgist or "geo-economic," "geo" referring to both land and George (Feder, 1996, p. 41).

As noted by Andelson (b, p. 386), "The doctrines of natural law and natural rights undergrid the entire framework of George's thought," although his positive-economic analysis is orthogonally independent and can stand alone without its normative complement. The natural-law philosopher John Locke

(1690) had argued that human beings are properly and equally self-owners. Taking the thought of John Locke to a logical conclusion, George argued that if one owns oneself, one owns one's labor, and so the taxation of labor and the products of labor is morally wrong.

On owning land, Locke stated that one could claim land by mixing one's labor with land, subject to the proviso that land of equal quality be left available freely to others. The implication seems to be that if such land is not available, then the title holders owe something to the rest of humanity, namely the rent, to be shared equally or spent for common purposes (Feder, 1996, p. 42).

Henry George made the moral case the core of his social philosophy, congruent with his economic analysis. He wrote (1883, p. 213), "It is no mere fiscal reform that I propose; it is a conforming of the most important social adjustments to natural laws." For George, there is a harmony between the economic and the moral: the policy of obtaining public revenue from land rent is both efficient and just. Moreover, when government provides the infrastructure, then it is just for the site owners who obtain the added rentals pay back those rentals to the government that generated them.

George's main interest was not so much public revenue but the prevention of what he considered to be the main social problem, poverty. By poverty George means not just the abject poverty of the homeless and of workers who can just barely pay for their subsistence, but also the middle-class deprivation of security, especially from the unemployment that occurs during depressions. His major work is therefore titled *Progress and Poverty* (1879), as an analysis of why poverty persisted in the midst of increasing technology and wealth.

George held that poverty is not due to any lack of natural resources, stating (1883, p. 78), "There is in nature no reason for poverty." George pinpointed the root cause of poverty in the land-tenure and tax systems. As Smith and other had recognized, much of the gains from economic progress adhere to site values—and we see an example of this in the rather high prices for real estate in places such as the San Francisco Bay Area, Tokyo, New York City, and other centers of commerce and industry. Much of this site value is generated by government-funded civic goods and services paid for mostly from the earnings of labor. Thus there is a huge forced transfer of wealth from workers to landowners, and the ownership of the most valuable commercial land tends to be highly concentrated in a few wealthy hands, as is the value of agricultural lands in many less-developed countries.

George stated that the landowners could retain some of the rent to facilitate the real-estate market. Thus the Georgist proposal to obtain all the rent for public revenue does not necessarily mean all of the rental, but rather the economic rent, that portion that is not needed to put the land factor to its most productive use.

2.3.   Land Theorists after Henry George

Leo Tolstoy was convinced by George's thought, and stated, "People do not argue with the teaching of George; they simply do not know it. He who becomes acquainted with it cannot but agree" (Chandler, 1982, p. 54). Whereas George saw the "single tax" as a means toward justice, Tolstoy also saw it as a transitional means to the achievement of an anarchist communal society, via its local administration Hecht (1997). Tolstoy (1906, p. 391) wrote, "Everyone has an equal right to the land and an inalienable right to the products of his labor … For the attainment of these rights only one means is necessary: the establishment of a single land tax."

Alfred Marshall retained the classical proposition of land rent as a surplus without an economic cost (Marshall, 1930, p. 156; Lackman, 1977). He also recognized that site rentals would increase with greater intensity of use. Marshall also objected to the neoclassical merging of land and capital goods, stating that "land must everywhere and always be classed as a thing by itself" (Marshall, 1961, Appendix G, pp. 802-803).

Knut Wicksell (1896 [1958], p. 113) wrote that "the general economic development of the community" increased the value of its land, and he proposed taxing such increases (p. 114). Friedrich von Wieser (1967 [1927], p. 340), an early theorist of urban rent, stated that "Urban rent is that part of the rental which is paid as a premium for the advantages of the better location," and these public-good advantages encompass the externalities present in the area. Léon Walras also favored taxing land rent (Cirillo, 1984) as a matter of distributive justice.

Franz Oppenheimer, author of *The State,* in which he traced the private ownership of land rent to conquest, also wrote an essay (1917) on land ownership which traced modern land law to the Roman Gracchi, which conferred the rent to the ruling landed class. From there the new Roman "quiritist" land tenure system was carried to Europe and the world. He urged that the Jewish settlers in Palestine (then under the Turkish empire) not copy the Roman system, and indeed the Jewish National Fund came to own most of the land in Israel, and lease it to residents, but at rents so low that the leaseholds amount to the full private ownership in practice, with the exception that the leasehold cannot be alienated to non-Jewish holders.

Harry Gunnison Brown was the main academic land-rent theorist of the first half of the 20th century. He studied under and assisted Irving Fisher, and was a Fisher monetarist as well as a follower of Henry George, holding that "the site value of land (which he considered unearned) should constitute the first source for governmental taxation" (Ryan, 1987, p. 82). His major work was *The Economics of Taxation,* a classic on the incidence of taxes. Among his conclusions was that a general sales tax ultimately falls on the owners of the factors of production. Brown held to the classical category of land as a distinct factor of

production, being nonreproducible and without a cost of production, with capital goods being a derivative factor (Brown, 1931, p. 273). Brown referred to himself "as an economist 'unemancipated' from the classical tradition" (Ryan, 1987, p. 83).

Paul Douglas (1972), famous along with Cobb for the Cobb-Douglas production function $Q = f(K^a L^b)$, which originated in Wicksteed, was favorably inclined to site-value taxation. Contemporary economists and other scholars who have written extensively on site-value and land-rent taxation include Kris Feder (1996), myself (Fred Foldvary), Mason Gaffney, Fred Harrison, Nicolaus Tideman, and the late William Vickrey. Of course many others have written on aspects of site-value taxation and the thought Henry George (see Blaug, 1992; Gaffney, 1982; Lissner and Lissner, 1991).

For Gaffney (1972), rent is the opportunity cost of land occupancy rather than simply the income from land. The site rental measures the annual amount which others would be willing to pay for its use, something the title holder cannot alter. Gaffney (1973) also contributed the proposition that since credit markets are imperfect due to uncertainty, people obtain credit on unequal terms, with the different discount rates not necessarily related to true risks. When most of the land rent is collected, the price of land falls, and thus the taxed rental substitutes for the present-value price that would otherwise have been paid. This "Gaffney effect" increases the efficiency of the land market.

Mason Gaffney also delved into the history of economic thought to uncover the reason why modern neoclassical economics, even while acknowledging the economic efficiency of taxing rent, pays so little attention to it in its macroeconomics and policy analysis. Gaffney (1994b) holds that there was a stratagem to recast economic theory, including its terminology, to evade the policy implications of Georgist theory. The neoclassical turn did not just adopt marginal utility theory but also folded land into capital, generalized the meaning of "rent," and relegated land to the minor specialty of "land economics."

Nevertheless, many contemporary and recent economists have agreed with Smith's proposition that ground rent is well suited for public revenue relative to other sources, without being full-scale advocates or writing much about it. Milton Friedman, for example, stated that "the least bad tax is the property tax on the unimproved value of land, the Henry George argument of many, many years ago" (Andelson, 1979a, p. 391).

In 1991, 30 economists, including three then Nobel-prize winners (one signer, William Vickrey, winning the prize later), signed a letter to Soviet president Mikhail Gorbachev advising him that "It is important that the rent of land be retained as a source of government revenue" (Tideman, 1991, p. 226). Had this prescription been heeded either by this last Soviet president or his Russian successor, along with secure and untaxed property rights to labor and investments, the massive capital flight and the financial crises in large part

due to tax evasion in Russia may well have been avoided. Nevertheless, this letter demonstrates that rent-based public finance continues to have adherents among economists of diverse backgrounds.

3.   ECONOMIC EFFECTS AND ETHICAL CONSIDERATIONS

3.1.   Excess Burdens

The "excess burden" of a tax is a social cost due to the reduction in the quantity of goods produced because the tax increases the price of the goods. It is "excess" because it is a cost beyond the resources that are shifted from private pockets to government coffers, but not lost to society (aside from their possibly being spent less efficiently). With an excess burden, resources do not get allocated to where people most want them, because the prices paid no longer reflect the marginal valuations by consumers and the marginal costs of the producers. Graphically, this excess burden, also called a deadweight loss and welfare loss, is the triangle between the supply and demand curves and the quantity after being reduced by the tax.

A sales or income tax or tax on real-estate improvements shifts up the supply curve, since the supply reflects the marginal costs of production, and these costs are not reduced just because the good is taxed. The supply curve shifts up along the demand curve to a higher price and lower quantity. In contrast, a tax on rent has no excess burden. The supply of land is fixed, and therefore there is no shift of supply. As explained above, a tax on rent does not affect the rent and also does not change the quantity of land nor the prices of production. The only change in price is that the price of land is capitalized down.

As noted above, a tax on rent can even have an excess benefit rather than burden if the actual supply of land is artificially reduced from the natural supply, such as by land held for purposes of prestige, consumption, or awaiting increasing rent and site values, when the rentals are subsidized by the financing of civic goods from non-rental sources. The Georgist tax shift to site rentals then moves the supply curve to the right towards the natural supply, and the rentals move down the demand curve to the new equilibrium, benefiting tenants, new buyers, and investors.

3.2.   Urban Sprawl and Blight

Urban blight can be defined as the reduction in the value of urban buildings and other improvements from what they would be in a pure free market. Slum areas have run-down buildings, capital goods that have not been well maintained. In a pure market, the owner has an incentive to keep his property well

maintained, as that maximizes the rental profit. Tenants will flee from poorly maintained housing.

But interventions change the incentives. Rent control implies that the landlord may not raise the rental to finance improvements if the owner may not pass on the full improvement costs. Legal delays and other costs imposed on evicting tenants also adds to the cost of landlording. Property taxes that fall on the whole property create perverse incentives in two ways. First, the improvement is penalized with additional taxes. Secondly, if the owner expects the area to be developed in the future, with civic improvements paid for by taxes other than on the site rentals, he can hold the site mainly for the anticipated future gain, and neglect the buildings. Site value thus reduces blight by eliminating the tax-punishment of improvements, and secondly by pushing site owners to put their land to its most productive current use, maximizing current rentals to pay the rental carrying cost, there being less gain from just holding the site.

Urban sprawl can be defined as a greater amount of land use than would be the case in a pure free market. Current urban policy subsidizes sprawl first with zoning and other restrictions that reduce the intensity of urban land use, and secondly with taxes that subsidize the urban fringes at the expense of the center. New developments require more civic infrastructure, schools, and other services, mostly paid for from taxes other than the site rentals.

Besides the distortions from taxes, the payments for utilities also typically subsidizes the suburbs. As Mason Gaffney (1964) has analyzed, utilities such as water are usually paid for by the amount used, regardless of location. But fringe users cost more than central-city users, as water to the fringe requires piping capacity all the way from the center. Gaffney also points out that urban sprawl does not just affect the city. Agricultural production is not just pushed further out, but various types of agriculture push others out, so that "urban sprawl sends out shock waves into the countryside which travel through the entire hierarchy of land uses" (p. 2).

Cities would become more compact with taxation based on site value and with utilities priced according to the cost of provision. Another requirement to bring cities to the pure market usage is the elimination of zoning that mandates maximum density and building codes that add to the cost of construction without providing benefits.

3.3.   Business Cycles

Henry George first theorized that real estate plays a key role in the business cycle, as speculative rises in site values make real estate too expensive to use for enterprise and investment. George was therefore also an early theorist of the role of expectations in cycle theory. Modern economics recognizes that economic investment, increases in the stock of capital goods, drives the business cycle. In the Austrian-school theory of the business cycle, artificially-low

interest rates due to excessive credit and monetary expansion distort invest-
ment, producing too many "higher order" long-term capital goods. The key
such capital good is real-estate construction, which amounts to about one quar-
ter or more of total investment (Matthews, 1967). Additionally, new houses
and offices require furniture and other durable goods. Much of real estate pur-
chases and construction is done with borrowed money, so that there is a finan-
cial dimension to the real-estate cycle as well. When real estate crashes, the
market prices of properties fall below the loan balance, the owners default, and
the fall in real estate then also brings the banks down. A Georgist-Austrian the-
ory of the business cycle thus joins together the real elements of capital goods
and land, and the financial elements of money, credit, and the rate of interest
(Foldvary, 1997).

Real-estate economist Homer Hoyt (1933), investigating the cycle of land
values in Chicago during the century prior to 1933, discovered a cycle of aver-
age duration of 18 years, coinciding with the major U.S. business cycle. Karl
Pribram (1940) recognized that increases in land values follow a rise in build-
ing activity, and that in the latter stages of a boom, real-estate costs render fur-
ther building unprofitable. As further analyzed by Fred Harrison (1983, pp. 64-
65), peaks in land values as well as in construction have preceded the major
depressions, indicating that the real-estate cycle is a likely cause of rather than
a consequence of the general business cycle, providing evidence consistent
with George's cycle theory. Harrison found similar cyclical relationships for
the UK and other countries. The dramatic rise and fall of real-estate in interwar
Germany was found by Bruno Heilig (1941) to be a key cause of the depres-
sion that led to Nazi rule. The boom and subsequent bust in Japan of the 1980s
and 1990s also followed the George-Hoyt-Harrison sequence of overbuilding,
high speculative land values, and then a collapse.

Aside from reforms in money and banking, tax reform that shifts public
revenue to site values reduces the land-value boom that can create excessive
construction in some areas, as investors anticipate that much of the gain will
come from increased site values, and also excessive buying and preparation of
undeveloped or underdeveloped sites, which later fail to become developed.
Tapping site rentals for public revenue and untaxing labor and capital thus
helps prevent the economic turbulence of periodic boom and bust cycles.

3.4.   The Ethics of Taxation

To Henry George, there is a harmony between economic efficiency and so-
cial equity. He wrote, "economic law and moral law are essentially one" (1879,
p. 560). Henry George regarded taxes on labor and capital goods as immoral
violations of self-ownership. Such taxes also impose added costs to production
that make such taxation problematical by raising costs above the economic

marginal costs. The benchmark for economic efficiency, price equal to marginal cost, is violated with any taxation that falls on enterprise. Thus, calls for economic justice that impose such taxes for redistribution are to that degree self-defeating in raising prices above marginal cost across the economic board, needlessly imposing an excess burden and social waste.

In Georgist ethics, human beings have an equal natural right both to their own labor and to an equal share of the benefits of the earth's natural endowment. This equality was also recognized by John Locke (1690). Since the right of self-ownership does not extend to what the self did not create, by default, the right to the equal benefit from nature's resources is equally held by all persons Foldvary (1999b). But "it is not necessary, in order to secure equal rights to land, to make an equal division of land. All that it is necessary to do is the collect the ground-rents for the common benefit" (George, 1883, p. 208). Moreover, the rentals created by civic goods provided by government may rightfully be created by government, otherwise this creates a subsidy to landowners at the expense of everyone else (Foldvary, 1999b).

4.   TERRITORY, GOVERNANCE, AND PUBLIC CHOICES

4.1.   Voting and Governance Structure

If governments base their principle public revenues from site rentals, the question then arises as to how such rentals are to be divided among the levels of government. The U.S. structure will be used here as a benchmark, with six levels of government: federal, State, county, township, neighborhood, and household. American Indian nations or tribes in reservation lands can be considered the equivalent of the State level. The "township" can be a municipality or a subdivision of a county.

One criterion for determining the distribution of the rental funds is the principle of subsidiarity, that a government service should be provided at the lowest level for which is it efficient in the broadest sense. An even more important principle for the community collection of the rental revenues is that the agency which generates the rentals should also be the one to collect them. A third principle adhered to by followers of Henry George is that the rent of land, shorn of all improvements both individual and civic, be spread equally among as wide a population as feasible.

Putting these principles together, the natural rent of land would be allocated to the households on an equal per-capita basis. The townships could then tap these rents for public revenue or leave them with the households, as determined by the contractual obligations of the residents. The rentals due to civic improvements would be collected by the level of governance or other agency which generate them. Subsidiarity would be implemented by placing services

and expenditures at the lowest level of government which can effectively provide them.

The assessment of the market site rentals can be separated, both institutionally as well as conceptually, from the collection and from the expenditure of the rentals. Subsidiarity implies that the collection take place no higher than at the county level. Real-estate taxes are now typically collected by county governments in the U.S. It is convenient for the property records to be local, so that the title holders can have access to them and to the officials responsible for them. Hence, the overhead involved in State or federal-level collection can be eliminated, since the main task is to send the owners bills and collect and record the funds. Since there is some fixed cost in maintaining the records, computers, and forms, there are probably economies of scale that usually place the efficient collection of the rental at the county rather than in the township.

The allocation of rentals to the level generating them would eliminate all top-down revenue sharing, since the funds would be allocated directly to that level of government. This would result in a more decentralized government provision of services than is now taking place, because to a substantial degree it is the revenue aspect that centralizes government in such services as education rather than the efficiency of provision. The efficiency in the broad sense includes the responsiveness of government to the desires of the public.

The assessment of the rentals, i.e., the estimation of the market prices, is not effectively done at a lower level. The local government would have an incentive to understate the rentals so that the local landowners pay less rental to the higher level governments. The federal government would have an incentive to overstate the rentals. A way to avoid such temptations would be to appoint local assessment boards made up of representatives from all levels of government.

In summary, a county-level board made up of representatives from all levels of government would supervise the assessments of the rentals. Some of the rentals would be retained by the county and some allocated to townships according to the estimates of rentals generated by their services. Some of the rental might also be allocated to the residents of the county on an equal per-capita basis. The remainder of the rental would flow up to the State government, which in turn would turn over a portion of its rental to the federal government, in proportion to the total site values in the States.

4.2.   Public Choice and Transfer Seeking

Public choice is the branch of economics that analyses choices that people make for a group rather than just for themselves. Regarding the choices of voters, a basic proposition of public choice is rational ignorance. An individual voter receives little material benefit from voting, and therefore has little

incentive, unless he has some personal interest, in extending much time and resources in obtaining information about the candidates and issues, or even voting.

In contrast, special interests for which the benefits are concentrated, each member obtaining a substantial gain, provide incentives for seeking transfers of wealth from taxpayers and consumers via governmental grants of subsidies. Such privileges, benefits arbitrarily granted to some and not others, can take various forms, including price controls, funds, and restrictions. This transfer seeking has been termed "rent seeking" because the pressure groups seek economic rents, funds beyond what would be obtained in a competitive market to put factors to their best productive use. A basic proposition in public choice theory is that transfer seeking takes place when recipients have concentrated benefits, and the costs are thinly spread out among the consumers and taxpayers, since members of the latter groups have little material incentive to expend resource to oppose the policy.

Such rent seeking explains why governments enact taxation with substantial excess burdens instead of minimizing that burden. The ownership of the most valuable sites, commercial and industrial areas, tends to be highly concentrated. A shift to taxing labor less and site rentals more would reduce the market value of their real estate. These owners have a strong incentive to finance candidates and ballot measures that avoid placing a tax or charge on site values and rentals. The typical voter is too ignorant to know about the burden-free alternative, and thus there is little political push for site-value revenues. Because of fiscal illusion based on ignorance of economics, there is less political resistance to taxing explicit flows, namely incomes and sales, than site values and rentals that are direct, sometimes implicit, and very visible.

The transfers to site values are largely implicit, as taxes on wages and capital returns finance infrastructure that elevates the site rentals. The tenants and general public regard the rentals as market based, not realizing that the implicit transfer of wealth skews these rentals. Such transfer seeking is also a function of the structure of voting. The governing paradigm is mass democracy, with thousands or millions of voters electing a candidate. This creates a demand for campaign funds, supplied by the special interests in return for favors. Legal limitations on campaign contributions attempt to treat the effects of the problem, usually with little effect, since the structure of government and voting create these powerful incentives that will find some way to be fulfilled.

The effective remedy for transfer seeking requires the removal of the cause, namely mass democracy. The logical alternative is small-group democracy, with voting divided into small cells, such as neighborhoods of a few hundred residents (Foldvary, 1999a; Foldvary, 2001). Little or no money is required to campaign for the local neighborhood council, leaving little demand for special-

interest financing. It is also much more costly for the interests to have to pro-
vide funds for thousands of small cellular elections.

The councils then from their members elect representatives to districts or
townships representing some number of neighborhood councils, and these in
turn elect higher-level councils, on to the federal level. Each election on each
level takes place in a small group, where electors can be personally acquainted
with the candidate and where it is not costly to be a candidate. The upward flow
of elections also creates an incentive to decentralize governance in accord with
subsidiarity, since the lower level can more closely monitor the next higher
level, and the power flows up from the bottom rather.

4.3.   Demand Revelation

The optimal amount of a public good that impacts territory was analyzed
above as that amount for which the marginal rental equals the marginal cost,
the total generated rental being greater than the total cost. In some cases, it
may not be known or even estimable how much the rental will rise. In that
case, the governing authority could ask the members or residents to state how
much the good is worth to them, i.e., the maximum they would be willing to
pay, which would then be the amount they would pay.

This would not necessarily reveal the true demand, since there would be
an incentive to hide one's preference to avoid paying. That can be remedied
by having each person pay the average cost of the good, and if anyone's stated
value changes the outcome relative to stating the average cost, then that person
also has to pay the social cost of his decision, the net differences between the
stated values and the average cost. Such a payment has been called a "Clarke
tax," after Edward Clarke, who discovered this demand-revealing method.

Demand-revelation with the Clarke payment is superior to conventional ma-
jority voting because it measures the intensity of preferences rather than giving
all preferences an equal share. The method also provides an incentive to reveal
one's true demand, since if one bids higher than one's true value to ensure
obtaining the good, one may have to pay a Clarke tax if that changes the out-
come, and understating one's value may result in not obtaining the item and
possibly changing the outcome to not getting it, and a Clarke payment, since
the good is obtained only if the total net values are positive.

For community services which do not impact on site rentals, demand reve-
lation with an equal per-capita payment can be the preferred method of fund-
ing. An example is assistance to an outside organization. Demand revelation
can also be combined with rental-based payments, first of for social decision
making, and then to partially finance the service with the per-capita payment
and partially with rental-based payments. The rentals could pay for the fixed-
cost investment, while the equal per-person payment pays the operating costs.

While a pure site-value charge might be more efficient in repaying the generated site rental, a modest per-capita charge would help reveal the effective demand for the civic good, hence be efficient in a broader sense.

5.   CONTRACTUAL COMMUNITIES AND PUBLIC RENTS

A private community is established by an explicit contract, in which each member voluntarily agrees to the terms, and which also has some procedure for terminating the contractual obligation. The community can be proprietary, or what Spencer MacCallum calls an "entrecom." A firm owns the site and the civic improvements, and the residents have leaseholds or rental agreements, possibly also owning the buildings they occupy. Examples include hotels, apartment houses, shopping centers, and land trusts.

The contractual community can also be an association of co-owners. Examples include condominiums, residential associations, and cooperative housing. Many of these provide streets, parks, recreation, security, and other civic services. The payments are usually a monthly rental or assessment. In some cases, they are equal payments per household, and in others they are based on the real-estate or site value. In many condominiums, each unit has a percentage interest that is inscribed in the master deed and may not be changed. The unit's voting and payment share are that percentage of the total, making the payment independent of the unit owner's own improvements (Foldvary, 1994).

While political pressures induce local governments to tax flows of income and sales, the market process induces contractual communities to use site rentals or some close approximation. They cannot piggyback on State and federal taxation, and the members expect the payments to be proportionate to benefits. If government enacted tax and service substitution, so that the private community could deduct from its tax obligations the funds that government saves by not providing theses goods, private communities could compete on an equal basis with governmental ones, and as communities are converted to contractual arrangements, so too would public finance shift from taxing transactions to tapping site rentals.

6.   CONCLUSION

Public revenue from land rent and site values has several advantages. There is little or no excess burden on the economy. Land cannot be hidden or go underground. The transaction costs of taxing site values are lower, with no need for audits or tax lawyers and accountants. It also avoids the implicit subsidies present when civic works become capitalized into land value. The reason that site values are not more widely tapped for public revenue is partly political, as the owners of commercial and industrial land would lose property value. But

most owners of residential as well as industrial land would have a substantial net gain if taxes shifted off of labor and capital and to site values. The elimination of the large excess burdens implies a social net gain much more than sufficient to compensate for any net losses. Thus the ultimate reason for the avoidance of public revenue from land rent is ignorance. Since private communities overcome such ignorance by following market incentives, a shift of civic services from governmental to private communities might be the most feasible way to achieve an efficient system of financing public goods.

REFERENCES

Andelson, Robert V. (ed.) (1979a). *Critics of Henry George.* London: Associated University Presses.

Andelson, Robert V. (1979b). "Neo-Georgism." In *Critics of Henry George,* Andelson, Robert V., ed. London: Associated University Presses. 381-393.

Andelson, Robert V. (ed.) (2000). *Land-Value Taxation Around the World,* 3rd ed. Annualsupplement to Volume 59, American Journal of Economics and Sociology and Blackwell Publishers.

Banks, Ronald. (ed.) (1989). *Costing the Earth.* London: Shepheard-Walwyn.

Barker, Charles Albro. (1955). *Henry George.* New York: Oxford University Press.

Blaug, Mark. (ed.) (1992). *Henry George (1839–1897).* Aldershot, U.K.: Edward Elgar Publishing.

Boudreaux, Donald and Pritchard, Adam. (1997). "Civil Forfeiture as a Tax." In *Taxing Choice: The Predatory Politics of Fiscal Discrimination,* pp. 347-367. London and New Brunswick: Transaction Publishers.

Brown, Harry Gunnison. (1917). "The Ethics of Land Value Taxation." *Journal of Political Economy* 25 (May): 464-492. Reprint Harry Gunnison Brown (1980). *Selected Articles by Harry Gunnison Brown: The Case for Land Value Taxation,* pp. 3-28. New York: Robert Schalkenbach Foundation.

Brown, Harry Gunnison. (1924). *The Economics of Taxation.* New York: Henry Holt. Reprint University of Chicago Press, 1979.

Brown, Harry Gunnison. (1931). *Economic Science and the Common Welfare.* Columbia, MO: Lucas Brothers.

Brown, Harry Gunnison. (1979) [1924]. *The Economics of Taxation.* Chicago: University of Chicago Press.

Brown, Harry Gunnison. (1980). *Selected Articles by Harry Gunnison Brown: The Case for Land Value Taxation.* New York: Robert Schalkenbach Foundation. [No editor indicated]

Chandler, Tertius. (1982). *The Tax We Need.* Berkeley: The Gutenberg Press.

Cirillo, Renato. (1984), "Economists and Social Reformers on Land Ownership and Economic Rents." *Revista Int. di Scienze Econ. e Comm* 31 (March): 3. Reprint *George and the Scholars,* Will and Dorothy Lissner, eds., pp. 209-224. NY: Roberty Schalkenbach.

Collier, Charles F. (1979). "Clark and Patten: Exemplars of the New American Professionalism." In *Critics of Henry George,* Andelson, Robert V., ed., pp. 261-272. London: Associated University Presses.

Cord, Steven. (1979). *Catalyst! How a Reform of the Property Tax can Revitalize our Cities and Counter Inflation and Recession.* Indiana, PA: Henry George Foundation of America.

Cord, Steven. (1985). "How Much Revenue Would a Full Land Value Tax Yield? Analysis of Census and Federal Reserve Data." *American Journal of Economics and Sociology* 44 (3): 279-293.

Cord, Steven. (1991). "Land Rent is 20% of U.S. National Income for 1986." *Incentive Taxation* July/August.

Davidson, J. Morrison. (1899). *[Concerning] Four Precursors of Henry George [and the single tax].* NY & London: Kennikat Press.

Douglas, Paul. (1972). *In the Fullness of Time.* New York: Harcourt Brace Jovanovich.

Dove, Patrick Edward. (1850) [1997]. *The Theory of Human Progression.* Reprint *An Anthology of Single Land Tax Thought,* Wenzer, Kenneth C., ed., 167-93. University of Rochester Press.

Dwyer, Terence M. (1982). "Henry George's Thought in Relation to Modern Economics." *American Journal of Economics and Sociology* 41 (4): 363-373. Reprint *George and the Scholars,* pp. 296-306. New York: Robert Schalkenbach Foundation, 1991. Reprint *Henry George (1839–1897),* pp. 132-142. Aldershot, UK: Edward Elgar Publishing.

Feder, Kris. (1996). "Geo-economics." In *Beyond Neoclassical Economics: Heterodox Approaches to Economic Theory,* Foldvary, Fred, ed., pp. 41-60. Aldershot, UK: Edward Elgar Publishing.

Foldvary, Fred E. (1980). *The Soul of Liberty.* San Francisco: The Gutenberg Press.

Foldvary, Fred E. (1989). "Rental Income in the USA: the Mystery of the Missing Billions." In *Costing the Earth,* Banks, Ronald, ed., pp. 177-181. London: Shepheard-Walwyn.

Foldvary, Fred E. (1994). *Public Goods and Private Communities.* Aldershot, UK: Edward Elgar Publishing.

Foldvary, Fred E. (1997). "The Business Cycle: A Georgist-Austrian Synthesis." *American Journal of Economics and Sociology* 56 (4): 521-541.

Foldvary, Fred E. (1998a). "Market-hampering Land Speculation: Fiscal and Monetary Origins and Remedies." *American Journal of Economics and Sociology* 57 (4): 615-637.

Foldvary, Fred E. (1998b). "Municipal Public Finance." In *Handbook of Public Finance,* Thompson, Fred and Green, Mark T., eds., pp. 397-446. New York: Marcel Dekker.

Foldvary, Fred E. (1999a). "Recalculating Consent." *Buchanan Festschrift,* <http://www.gmu.edu/jbc/fest/files/foldvary.htm>

Foldvary, Fred E. (1999b). "The Ethics of Rent." In *Land-Value Taxation: The Equitable and Efficient Source of Public Finance,* pp. 184-204. NY: M.E. Sharpe, and London: Shepheard-Walwyn.

Foldvary, Fred E. (2001). "The Completely Decentralized City: The Case for Benefits Based Public Finance." *American Journal of Economics and Sociology* 60 (1): 403-418.

Gaffney, Mason. (1964). *Containment Policies for Urban Sprawl.* Lawrence KS: University of Kansas Publications, Governmental Research Series No. 27.

Gaffney, Mason. (1968). "The Valuation of Public Goods." In *Social Sciences and the Environment,* Garnsey, Morris and Hibbs, James, eds., pp. 154-160. Boulder: University of Colorado Press.

Gaffney, Mason. (1970). "Adequacy of Land as a Tax Base." In *The Assessment of Land Value,* Holland, Daniel, ed. Madison: Univ. of Wisconsin Press.

Gaffney, Mason. (1972). "Land Rent, Taxation, and Public Policy: The Sources, Nature, and Functions of Urban Land Rent." *American Journal of Economics and Sociology* 31: 241-257.

Gaffney, Mason. (1973). "Tax Reform to Release Land." In *Modernizing Urban Land Policy,* Clawson, Marion, ed. Baltimore: Johns Hopkins Press.

Gaffney, Mason. (1976). "Toward Full Employment with Limited Land and Capital." In *Property Taxation, Land Use, and Public Policy,* Lynn, Arthur D. Jr., ed. Madison: University of Wisconsin Press.

Gaffney, Mason. (1982). "Two Centuries of Economic Thought on Taxation of Land Rents." In *Land Value Taxation: The Progress and Poverty Centenary,* Lindholm, Richard W. and Lynn, Arthur D. Jr., eds., pp. 151-195. Madison: University of Wisconsin Press.

Gaffney, Mason. (1989). "The Role of Ground Rent in Urban Decay and Revival." *Distinguished Papers* 89F-1. Business Research Institute, St. Johns University.

Gaffney, Mason. (1994a). "Land as a Distinctive Factor of Production." In *Land and Taxation,* Tideman, N., ed., pp. 39-102. London: Shepheard-Walwyn.

Gaffney, Mason. (1994b). "Neo-Classical Economics as a Stratagem against Henry George." In *The Corruption of Economics,* Gaffney, Mason and Harrrison, Fred, eds., pp. 29-163. London: Shepheard-Walwyn.

Gaffney, Mason, and Fred Harrison, (eds.) *The Corruption of Economics.* London: Shepheard-Walwyn.

George, Henry. (1879) [1975]. *Progress and Poverty.* New York: Robert Schalkenbach.

George, Henry. (1883) [1966]. *Social Problems.* New York: Robert Schalkenbach.

George, Henry. (1886) [1980]. *Protection or Free Trade.* New York: Robert Schalkenbach.

George, Henry. (1890). "The Single Tax: What it Is and Why We Urge It." *The Christian Advocate.* Reprinted by the Robert Schalkenbach Foundation, New York, NY.

George, Henry. (1894) [1981]. *The Science of Political Economy.* New York: Robert Schalkenbach Foundation.

Hardin, Garrett. (1968). "The Tragedy of the Commons." *Science* 162: 1243-1248.

Harriss, C. Lowell. (ed.) (1973). *Government Spending and Land Values.* Madison: University of Wisconsin Press, Reprint No. 57.

Harriss, C. Lowell. (1976). *Local Responsibility and Land Taxation: Lessons from the United States Experience.* Singapore University Press. Reprint Washington: American Enterprise Institute.

Harrison, Fred. (1983). *The Power in the Land.* New York: Universe Books; London: Shepheard-Walwyn.

Harrison, Fred. (ed.) (1998). *The Losses of Nations.* London: Othila Press.

Heath, Spencer. (1936). *Politics Versus Proprietorship.* Manuscript.

Heath, Spencer. (1957). *Citadel, Market and Altar.* Baltimore: Science of Society Foundation.

Hecht, Leo. (1997). "Tolstoy on Land, Possessions, Work, and the Single Tax." In *An Anthology of Single Land Tax Thought,* Wenzer, Kenneth C., ed., pp. 377-402. Rochester: University of Rochester Press.

Heilig, Bruno. (1941). "Why the German Republic Fell." In *Why the German Republic Fell and Other Studies of the Causes and Consequences of Economic Inequality,* Madsen, A. W., ed. London: Hogarth Press. Reprint New York: Robert Schalkenbach Foundation.

Holterman. (1972).

Hoyt, Homer. (1933). *One Hundred Years of Land Values in Chicago.* Chicago: University of Chicago Press.

Lackman, Conway. (1977). "The Modern Development of Classical Rent Theory." *American Journal of Economics and Sociology* 36 (1). Reprint *George and the Scholars,* pp. 239-251. New York: Robert Schalkenbach Foundation, 1991.

Lindholm, Richard W., and Arthur D. Lynn, Jr. (1982). *Land Value Taxation.* Madison: University of Wisconsin Press.

Lissner, Will and Lissner, Dorothy. (eds.) (1991). *George and the Scholars.* New York: Robert Schalkenbach Foundation.

Locke, John. (1690) [1947]. *Two Treatises of Government.* New York: Hafner Press.

MacCallum, Spencer. (1965). "The Social Nature of Ownership." *Modern Age* 9 (1) (Winter 1964/5): 49-61.

MacCallum, Spencer. (1970). *The Art of Community.* Menlo Park: Institute for Humane Studies.

MacCallum, Spencer. (1971). "Jural Behavior in American Shopping Centers: Initial Views on the Proprietary Community." *Human Organization* 30 (1): 3-10.

MacCallum, Spencer. (1972). "Associated Individualism: a Victorian dream of freedom." *Reason* 4 (1): 17-24.

MacCallum, Spencer. (1986). "Outgrowing Government as We Know It." From a lecture. San Pedro, CA: The Heather Foundation.

Marshall, Alfred. (1930). *Principles of Economics,* Vol. I. 7th ed. London: Macmillan.

Marshall, Alfred. (1961). *Principles of Economics,* 9th ed. London: C. W. Guillebaud.

Matthews, R. C. (1967). *The Business Cycle.* NY: Cambridge University Press.

Mieszkowski, Peter and Zodrow, George R. (1989). "Taxation and the Tiebout Model." *Journal of Economic Literature* 27 (3): 1098-1146.

Miles, Mike. (1990). "What is the Value of All U.S. Real Estate?" *Real Estate Review* 20 (2): 69-75.

Mill, John Stuart. (1908). *Principles of Political Economy,* Vol. I, 7th ed. London: Cooperative Publishing Society.

Musgrave, R. A. and Peacock, Alan T. (eds.) (1958). *Classics in the Theory of Public Finance* (1958). London: Macmillan & Co.

Noyes, Richard. (ed.) (1991). *Now the Synthesis.* London: Shepheard-Walwyn; New York: Homes & Maier.

Oates, Wallace E. (1969). "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis." *Journal of Political Economy* 77 (6): 957-971.

Oppenheimer, Franz. (1917). "Collective and Private Ownership of Land." In *Land Tenure in Palestine.* The Hague: Jewish National Fund. Reprint Wenzer, Kenneth C. (1997). *An Anthology of Single Land Tax Thought,* pp. 414-434. Rochester: University of Rochester Press.

Paine, Thomas. (1797). *Agrarian Justice.* Reprint Wenzer, Kenneth C., ed. 1997. *An Anthology of Single Land Tax Thought,* 143-156. Rochester: University of Rochester Press.

Peterson, V. G. and Tseng Hsiao. (2000). "Kiao-chau." In *Land-Value Taxation Around the World.* Andelson, Robert V., ed., pp. 365-369. *American Journal of Economics and Sociology,* supplement 59.

Pribram, Karl. (1940). "Residual, Differential, and Absolute Urban Ground Rents and their Cyclical Fluctuations." *Econometrica* 8 (January): 62-78.

Pullen, John. (2001). "Henry George's Land Reform: The Distinction between PrivateOwnership and Private Possession." *American Journal of Economics and Sociology* 60 (2): 547-56.

Ricardo, David. (1821) [1984]. *The Principles of Political Economy and Taxation.* Reprint London and Melbourne: J. M. Dent. & Sons.

Richman, Raymond L. (1973). "The Capitalization of Property Taxes and Subsidies." In *Government Spending and Land Values,* Harriss, C. Lowell, ed., pp. 19-27. Madison: University of Wisconsin Press.

Ryan, Christopher K. (1987). *Harry Gunnison Brown, Economist.* Boulder: Westview Press.

Sakolski, Aaron M. (1957). *Land Tenure and Land Taxation in America.* New York: Robert Schalkenbach Foundation.

Saunders, Vernon I. (1999). "In Defense of the Two-Rate Property Tax." In *Land-ValueTaxation: The Equitable and Efficient Source of Public Finance,* pp. 269-276. NY: M.E. Sharpe and London: Shepheard-Walwyn.

Schwartzman, Jack. (1997). "Thomas Paine and the Ground-Rent Proposal." In *An Anthology of Single Land Tax Thought,* Wenzer, Kenneth C., ed., pp. 133-158. Rochester: University of Rochester Press.

Silagi, Michael. (1984). "Land Reform in Kiaochow, China: From 1898 to 1914 the Menace of Disastrous Land Speculation was Averted by Taxation." *American Journal of Economics and Sociology* 43 (2): 167-77.

Smith, Adam. (1976a) [1776]. *The Wealth of Nations,* Vol. 1, Canaan, Edwin, ed. Chicago: University of Chicago Press.

Smith, Adam. (1976b) [1776]. *The Wealth of Nations,* Vol. 2, Canaan, Edwin, ed. Chicago: University of Chicago Press.

Spence, Thomas. (1775) [1997]. "The Rights of Man." In *An Anthology of Single Land Tax Thought,* Wenzer, Kenneth C., ed., pp. 54-69. Rochester: University of Rochester Press.

Spencer, Herbert. (1850) [1995]. Social Statics. New York: Robert Schalkenbach Foundation.

Stiglitz, Joseph E. (1977). "The Theory of Local Public Goods." In *Economics of Public Services:* 274-333.

Stiglitz, Joseph E. (1983). "The Theory of Local Public Goods Twenty-Five Years after Tiebout: A Perspective." In *Local Provision of Public Services: The Tiebout Model after Twenty-Five Years,* Zodrow, George R., ed., pp. 17-54. New York: Academic Press.

Stiglitz, Joseph E. (1986). *Economics of the Public Sector.* New York: W. W. Norton.

Tideman, T. Nicolaus. (1985). "Efficient Local Public Goods Without Compulsory Taxes." In *Perspectives on Local Public Finance and Public Policy* 2: 181-202.

Tideman, T. Nicolaus. (1990). "Integrating Land-Value Taxation with the Internalization of Spatial Externalities." *Land Economics* 66 (3): 341-355.

Tideman, T. Nicolaus. (1991). "Open Letter to Mikhail Gorbachev." In *Now the Synthesis,* Tideman, T. Nicolaus, ed., pp. 225-230. London: Shepheard-Walwyn; New York: Homes & Maier.

Tideman, T. Nicolaus. (ed.) (1994a). *Land and Taxation.* London: Shepheard-Walwyn.

Tideman, T. Nicolaus. (1994b). "The Economics of Efficient Taxes on Land." In *Land and Taxation,* Tideman, T. Nicolaus, ed., pp. 103-140. London: Shepheard-Walwyn.

Tideman, T. Nicolaus and Plassmann, Florenz. (1998). "Taxed Out of Work and Wealth: The Costs of Taxing Labor and Capital." In *The Losses of Nations,* Harrison, Fred, ed., pp. 146-174. London: Othila Press.

Tideman, T. Nicolaus. (1999). "Taxing Land Is Better than Neutral: Land Taxes, Land Speculation, and the Timing of Development." In *Land-Value Taxation: The Equitable and Efficient Source of Public Finance,* Wenzer, Kenneth C., ed., pp. 109-133. NY: M.E. Sharpe, and London: Shepheard-Walwyn.

Tideman, T. Nicolaus, and Tullock, Gordon. (1976). "A New and Superior Process for Making Social Choices." *Journal of Political Economy* 84(6): 1145-1159.

Tiebout, Charles M. (1956). "A Pure Theory of Local Expenditure." *Journal of Political Economy* 64: 416-424.

Tolstoy, Lev. (1906). "The Only Possible Solution to the Land Question." In *An Anthology of Single Land Tax Thought,* Wenzer, Kenneth C., ed., pp. 389-394. Rochester: University of Rochester Press.

Tucker. (1958).

Velde, Françoise R. (1997). "The Physiocrats and the Single Tax." In *An Anthology of Single Land Tax Thought,* Wenzer, Kenneth C., ed., pp. 13-23. Rochester: University of Rochester Press.

Vickrey, William. (1977). "The City as a Firm." In *Economics of Public Services,* Feldstein, Martin and Inman, Robert, eds., pp. 334-343. Reprint Vickrey, William. (1994). *Public Economics,* Arnott, Richard et al., eds., pp. 339-349. Cambridge, U.K.: Cambridge University Press.

von Wieser, Friedrich. (1967) [1927].

Walras, Léon. (1896). *Studies in Social Economics.* Lausanne: F. Rouge and Co. Unpublished trans. by Mason Gaffney, 1967.

Wenzer, Kenneth C. (ed.) (1997). *An Anthology of Single Land Tax Thought.* Rochester: University of Rochester Press.

Wenzer, Kenneth C. (ed.) (1999). *Land-Value Taxation: The Equitable and Efficient Source of Public Finance.* NY: ME. Sharpe, and London: Shepheard-Walwyn.

Wicksell, Knut. (1958). "A New Principle of Just Taxation." In *Classics in the Theory of Public Finance,* Musgrave, R. A. and Peacock, Alan T., ed., pp. 72-116. London: Macmillan & Co. Reprint in part in *Public Choice and Constitutional Economics,* Gwartney, James D. and Wagner, Richard E., eds., pp. 117-130. London and Greenwich, CT: Jai Press, 1988.

Chapter 8

DEBT, MONEY, AND PUBLIC FINANCE

Richard E. Wagner
George Mason University
rwagner@gmu.edu

Abstract   This chapter starts by reviewing various arguments about the possible inter-temporal shifting of the burden of public debt. Within a democratic polity, it is misleading to speak of the state as being indebted, even though the historical record presents us with many instances of indebted monarchs. A democratic state is rather in the position of a financial intermediary, though the relationships among debtors and creditors are governed by political and not market institutions. Under prevailing monetary institutions, moreover, public debt is integrated into contemporary arrangements concerning money and credit.

Keywords: Public debt burden, inter-temporal shifting, Ricardian equivalence, seigniorage, rent seeking

JEL classification: D72, E42, E60, H63

A government can finance its activities in four ways. It can tax. It can borrow. It can create money. And it can generate revenue through its operation of enterprises.[1] These days, taxation is widely regarded as the primary source of state revenue, and this primacy is generally accepted as proper. It is quite common for contemporary authors to cite Adam Smith's four canons of taxation, which he articulated in 1776 in the *Wealth of Nations.* These canons held that taxes should be levied in proportion to property, should be certain and not arbitrary, should be convenient to pay, and should be economical to administer for both the taxpayer and the state. Furthermore, Smith thought that taxation ideally should be the sole source of state revenue. He preceded his discussion of tax canons with an argument that the state should abolish its holdings of property, thereby relinquishing any revenue it derives from those holdings. Modern states, of course, have not followed Smith's advice in this respect, and have proven ready to accept revenue from nearly any source.

While Smith's vision of a state financed predominately by taxation is second nature to contemporary fiscal scholars, it has not always been this way, either in theory or in practice. In 1760, Johann Heinrich Gottlob von Justi articulated a quite different vision of state finance in *Natur und Wesen der Staaten.* Taxation was a last resort instrument of public finance. For Justi and the Cameralist authors generally, states were to be financed in the first instance through revenues generated from state enterprises and lands. Justi argued that ideally states would not tax at all, and would derive all of their revenue from their enterprises and lands. Taxation was a secondary option only, and one, moreover, that was more strictly limited in Justi's canons than in Smith's. Justi's canons covered all the territory covered by Smith's canons, and more. Justi also held that a tax should never deprive a taxpayer of necessaries or cause him to reduce his capital to pay the tax, nor should a tax ever harm the welfare of taxpayers or violate their civil liberties. This Cameralist principle of public finance, moreover, received practical implementation throughout the central European lands where the cameralists were influential, as illustrated by the much greater fiscal significance of enterprise revenues in the cameralist lands, where those revenues generally provided the majority of state revenues.[2]

This difference between Justi and Smith reflects one of the important orienting principles of the cameralists, namely, that the state acts as a participant within the society and its economic order. The cameralist advice on the use of state budgets and other policy instruments to promote the well being of the state and its subjects took place within a presumption that the state itself was located inside the economic order and not outside it. The state is but another participant within the economic order of a society. Civil society and the state are nonseparable and co-emergent. This treatment of the state in relation to civil society contrasts sharply with various contemporary constructions where state and society are treated as autonomous and independent from each other. In this alternative construction, the state intervenes into civil society and its processes. This distinction between the state as participating within the economic order and the state as intervening into the economic order, has numerous implications and ramifications, one of which concerns the generation of state revenues. The cameralist ideal, recognizing that practice rarely if ever conforms fully to ideals, was the state as a peaceful and productive participant within the economic order. The Smithian ideal was the state as a violent force for intervention into the economic order. It is perhaps no wonder that Joseph Schumpeter (1954, p. 172) described Justi as "A. Smith … with the nonsense left out."

Geoffrey Brennan and James Buchanan (1980) construe the state as a revenue-maximizing beast, a leviathan. While the leviathan of the Bible lived in the sea, it is easy enough to imagine it as living on the land. Smith's maxims for taxation are a recipe for living with the leviathan by doing such things as

clipping the beast's nails and filing down its teeth, A beast it will always be, and the objective of tax maxims should be to limit the damage the beast causes. Justi's maxims for taxation, in conjunction with his preference for enterprise revenues over taxation, represent a contrary intellectual orientation that would seek to domesticate the beast. To be sure, some would argue that genuine domestication is impossible.

Regardless of the relative standing of taxes and enterprises as sources of state revenue, fiscal scholars have generally regarded borrowing and money creation as secondary forms of public finance. Indeed, borrowing and creating money have often been characterized as instruments of *extraordinary* public finance, in contrast to taxes (and, once upon a time, enterprise revenues) as being instruments of *ordinary* public finance. Borrowing and creating money are not options different from taxation, but are different forms of taxation. A state that borrows is reducing current tax extractions in exchange for making a commitment to impose higher tax extractions in the future to service and amortize the debt. Borrowing is simply deferred taxation. Money creation is also a form of taxation, though one that is collected currently and not in the future. A state could impose a tax directly on money. Such a tax, however, would be costly to implement and enforce. It is cheaper for a state to tax money indirectly by debasing its real value through inflating the supply of money.

This essay explores various issues and controversies regarding borrowing and money creation as instruments of public finance. It starts by reviewing a controversy about public debt that that started shortly after the end of World War II. This controversy centered on the ability of public debt to transfer the burden of current state spending from current taxpayers to future taxpayers. A review of this controversy will help to set up the subsequent examination of debt and money as instruments of public finance. This analysis will pay particular attention to the institutional framework within which governments are constituted and borrowing occurs. It matters a great deal analytically whether governments are autocratic or democratic, as well as whether there is freely competitive banking or state-imposed central banking.

1. POSTWAR CONTROVERSY OVER THE BURDEN OF PUBLIC DEBT

The postwar period has seen a sequence of controversies about the locus of the burden of government expenditures that are financed by public debt. The classical theory of public debt held that state borrowing allowed the cost of current public spending to be shifted forward to the time when the debt was amortized. A good deal of argument took place over the propriety of such an intertemporal shift in tax burden. If future taxpayers are not represented when public debt is chosen, present taxpayers might choose too much public

debt because this shifts some of the burden of public spending from present to future taxpayers. On the other hand, current public spending might provide significant benefits for future taxpayers. In this case, public debt could allow tax burdens to be placed more fully on the beneficiaries of public spending. The use of public debt to finance wars, for instance, was generally thought to be desirable because it allowed the present generation, which fought the war, to shift some of the burden onto future generations, which received benefits from the battles fought by the current generation. In contrast, the use of public debt to finance government consumption was opposed on the grounds that it would allow a current generation to shift the burden of their own state-provided enjoyment onto future generations.

While the classical theory allowed for public debt, it also placed strong limits on its use, due to the perceived opportunities for the exploitation of future taxpayers by current taxpayers. The classical theory of public debt was challenged by the Keynesian formulations that were ascendant in the early postwar period. The classic statement of the Keynesian position was set forth in Abba Lerner (1948). There was a straightforward denial that public debt could shift the burden of public expenditure forward in time, thereby undercutting the classical warning against public debt. Whether a war is financed by current taxation or by borrowing, the resources expended to fight the war are sacrificed in the present and not in the future. Indeed, it makes no difference whether public debt is used to finance public capital creation or current consumption. In either case, the cost of public spending is necessarily borne in the present when it is incurred. The subsequent servicing of national debt will require the imposition of taxes to pay interest and to amortize the debt. Those tax payments, however, do not represent any bearing of cost stemming from past decisions, for that cost was borne in the past when the expenditures were made.

Present taxes to cover interest on and amortization of the debt are simply transfer payments, as distinct from being genuine opportunity costs. Today's public debt entailed a burden yesterday when it was created, but the service of that debt today is only a transfer payment and not a genuine burden. If public debt involves any current burden, it must be due to some secondary effect of the debt in reducing the stock of capital from what it might otherwise be. This effect was generally thought to be conceivable to the extent that government borrowing led to increased interest rates. To the extent that such an effect might operate, public debt could involve a secondary burden upon present taxpayers. The primary burden, however, was borne in the past, when the debt-financed expenditures were made. The only exception to this proposition about the primary burden of public debt would arise to the extent that public debt was held externally by citizens of foreign lands. To the extent public debt was held internally by domestic nationals, however, public debt places no burden on future

taxpayers, and debt service is simply a transfer payment between bondholders and future taxpayers.

This Keynesian rejection of the classical, cost-shifting thesis about public debt was countered sharply in James Buchanan (1958). Buchanan advanced an alternative articulation of the classical argument that debt finance allows the burden of public expenditure to be shifted forward in time. Buchanan, moreover, was referring to the primary burden, independently of any secondary impacts that might operate through capital formation. He further argued that the distinction between internal and external debt was irrelevant because the same implication for cost shifting resulted in either case.

Buchanan did not dispute the Keynesian point that it is impossible to shift resources from the future to the present. The resources that are used to supply governmental activities currently must be taken from what would otherwise be available currently for private use. But this does not mean that borrowing and taxing have the same temporal location of cost. With taxation, costs are clearly borne in the present, just as costs are borne in the present when someone pays cash rather than borrows. But public borrowing creates a strikingly different situation. If a war is financed by public borrowing, the people who buy the bonds provide the revenues required to fight the war. Yet these bondholders do not bear any burden. To the contrary, they have willingly reduced their current consumption to provide the resources for government to fight the war, in exchange for even greater consumption in the future when government services and amortizes the debt. Bondholders have been induced voluntarily to shift their intertemporal pattern of consumption, and have thereby secured a preferred state of affairs, and have most certainly not borne any burden. Present taxpayers obviously bear no burden, because the issue of public debt has reduced their taxes from what they would otherwise have been. The burden of the current spending must thus be borne by future taxpayers who must pay to amortize the debt, when they would not have had to make such payments had tax finance been used initially.

Public debt involves a two-part transaction and not just a single transaction between citizens and the state. In one part of the public debt transaction, bondholders agree to pay taxes that would otherwise have been paid by taxpayers. In the other part of the transaction, taxpayers in the future are being committed now to making payments to the bondholders. It is the bondholders when the debt was created who provide the resources necessary to supply the public output in question. But these bondholders bear no burden because they have been compensated for this through the promise of future payments for interest and amortization. And certainly taxpayers in the present cannot have borne any burden, because they secured tax reductions by virtue of the purchase of bonds by bondholders. The burden of debt-financed spending must

rest on future taxpayers, which implies that public debt does shift the cost of public spending forward in time.

Just as a Keynes vs. the Classics kind of controversy raged in macro and monetary economics in the postwar period, so did a Keynes vs. the classics kind of controversy rage over public debt. A good deal of this controversy is captured in James Ferguson's (1964) collection of essays. Just as the macro and monetary controversies took on a new character with the development of rational expectations and New Classical macroeconomics, so too did the public debt controversies. The sharpest statement of this position was set forth in Robert Barro (1974), though Earl Thompson (1967) articulated a similar position. The New Classical position embraced the classical claim set forth by Buchanan, that public debt allows the burden of public expenditure to be shifted forward in time. Contrary to Buchanan and the classics, though, Barro held that this shifting was of little consequence because of David Ricardo's (1817) proposition about the equivalence of debt and taxes.

For a single person, Ricardian Equivalence is little more than simple arithmetic. One person may pay a $100 tax obligation now. Another might pay it by borrowing $100 for one year at ten percent. This person's payment of $110 in one year is equivalent to the other person's present payment of $100. Whichever method of payment is selected, the taxpayer's net worth remains the same. As a simple matter of arithmetic, borrowing and taxing are equivalent in present value terms.

Whether they are also equivalent for the theory of public finance is another matter. Ricardo himself thought that the fiscal effects of debt would differ from those of taxation, with debt promoting public profligacy. Whether the simple arithmetic of Ricardian Equivalence applies politically as well as individually has created considerable controversy. If Ricardian Equivalence does hold in a political setting, public debt will not cause present taxpayers to think they are wealthier, because they will take into account the present value of the future tax liabilities that will be necessary to service and amortize the debt. Contrary to Buchanan and the classics, public debt would raise no normative issues of intergenerational equity. Contrary to the Keynesians, the government's budget could not serve as a tool to promote economic stability because an increase in state borrowing would not expand aggregate spending. Public debt would be neither an evil nor a gallant animal, as it rather was simply neutered in the New Classical formulation.

To say that public debt was neutered in one particular analytical formulation is not, of course, to say that it was neutered in reality. The New Classical formulation treats public debt as a form of personal debt and, indeed, treats the state simply as a wealthy individual. The New Classical formulations might thus seem particularly applicable to the conduct of autocratic regimes where state policies are the choices of the autocrat. In democratic regimes, however,

state policies are not the choices of some ruler. Rather, they emerge through complex interactions among political participants, and with those interactions being shaped and constrained by some particular institutional framework that governs the relationships among those participants. Despite the universal validity of the arithmetic of Ricardian Equivalence, the fiscal economics of public debt may differ, depending on the particular framework of political and fiscal institutions that are in place.[3] A shift from taxation to debt would thus be more a micro than a macro matter, and in this respect would be treated as but one instance of a shift in the structure of taxation. Public debt, as taxation generally, emerges out of a budgetary process. The central issues about public debt are of the same nature as all other budgetary controversies within a society. Public debt must be located within models of budgetary politics, and when this is done public debt becomes centrally concerned with interest groups and wealth redistribution among current taxpayers. Any intertemporal redistribution becomes incidental to the primary redistribution that lies at the core of the budgetary process.[4]

2. AUTOCRACY AND THE PERSONALIZATION OF PUBLIC DEBT

Public debt analysis mostly adopts autocracy as its ideal-type state, even if the analysis is not expressed this manner. Such an ideal typical state is implicit in the very treatment of public debt as simply a special case of private debt. To be sure, in the autocratic states of the ancient regime, state accounts and personal accounts were mingled. A state's debts were the debts of a king or prince, and were perhaps a source of concern for creditors as well. All public debt was necessarily held externally, for it would make so sense for a king or prince to borrow from himself. A state could be a net debtor or it could be a net creditor. In either case, though, it would be reasonable to treat the state as a person, though often a particularly large and powerful one. A king's borrowing to fight a war would be analytically indistinguishable from an individual's borrowing to buy a car.

Autocracy provides the simplest institutional framework for the consideration of public debt. There is no significant difference, conceptually speaking, between public and private debt in this case. To be sure, the fiscal literature often describes public debt as "sovereign debt." This description conveys the idea that public debt is just a particular form of private debt, where the borrower is especially powerful, which in turn may create some unique situations of borrower-lender relationships that do not arise in typical credit transactions.

Just as a private citizen might borrow in response to some unanticipated decline in income or some unanticipated increase in desired spending, so might

a sovereign. For a sovereign, the unanticipated decline in income might correspond to a depression. The unanticipated increase in desired spending might correspond to a desire to go to war. In either case, borrowing is the alternative to an increase in taxes to cover these unanticipated changes in income or expenditure. State borrowing allows a sovereign to achieve a smoother intertemporal pattern of taxation than would be possible otherwise (Barro, 1979). There are several possible reasons why an autocrat might be interested in the tax smoothing that borrowing allows. The excess burden of a tax generally rises with the square of the rate of tax. Borrowing and tax smoothing thus leads to a lower excess burden in the aggregate than would result under a regime of a continually balanced budget. A very benevolent autocrat might borrow because he valued the increased welfare of his subjects that resulted from this reduction in excess burden. A non-benevolent autocrat might borrow because this reduction in excess burden lowered the resistance to taxation among his subjects, under the presumption that the intensity of tax resistance varied directly with both the excess burden and the amount of taxes extracted.

For an autocrat, as for an individual citizen, public debt affects the timing of expenditure but does not affect net worth. Ricardian Equivalence clearly holds for personal debt. Someone who borrows to buy a car does not become wealthier than he would have been by paying cash. The reduction in cash that would have been required is equivalent to the present value of the liability for amortization payments when the car is financed by borrowing. Loan finance may be preferred to cash finance in some cases, perhaps as illustrated by arguments that loan finance allows the buyer to achieve a smoother time path of total consumption than would be possible with cash finance. But loan finance does not allow someone to escape from the intertemporal budget constraint; consumption may be reduced by a lesser amount now with loan finance, but it will have to be reduced by a greater amount in the future as the debt is amortized.

The Ricardian character of personal loan finance stems from the institutional setting within which credit markets operate, namely a framework of free exchange organized within an institutional framework of property and contract.[5] Within this institutional framework, a borrower cannot expect his borrowing to increase his net worth. A borrower who held such an expectation would be receiving a gift and not a loan. Lenders won't lend if they believe this will simply transfer their wealth to borrowers. Borrowers must convince lenders to lend, which lenders will do only if they feel confident that those loans will be paid. Such institutional practices as credit references and collateral, along with the readiness of the state to enforce credit contracts, generates Ricardian Equivalence as a feature of ordinary credit markets.

A sovereign, however, is not an ordinary participant in a credit market. The sovereign's debt is a personal debt, just as fully as are the debts of other borrowers. The institutional framework is altered when sovereign debt is involved.

For one thing, the sovereign cannot be called upon to enforce debt contracts against himself. A lender cannot call upon some external authority to enforce a contract against a sovereign. So long as debt is supplied voluntarily to the sovereign, the sovereign must have generated an expectation of contractual compliance among the lenders. A good deal of the recent literature on political economy has explored methods of commitment and concerns with reputation as a means of generating such expectations among lenders (see, for instance, Bulow and Rogoff (1989), Calvo (1988), Drazen (2000, pp. 101-215), Eaton and Gersovitz (1981), and Grossman and Van Huyck (1988)).

Moreover, contracts of sovereign debt need not be genuinely voluntarily. They can be duressful instead. Someone with wealth may well prefer not to lend to a sovereign, and yet do so anyway in light of what might be the consequences of refusing to lend. In ordinary credit markets, lenders are voluntary creditors. With sovereign debt, however, lenders may well be forced creditors, with a veneer of voluntarism masking the coerced reality that surrounds the loan. Under these circumstances, both borrower and lender may expect the transaction to increase the sovereign's net worth and to reduce the lender's net worth.

3.   DEMOCRACY AND THE INTERMEDIARY STATE

In his famous essay on just taxation, Knut Wicksell (1896, p. 82) complained that the theory of public finance "still rests on the now outdated political philosophy of absolutism. The theory seems to have retained the assumptions of its infancy, in the seventeenth and eighteenth centuries, when absolute power ruled almost all Europe." While the theory of public finance has changed greatly since Wicksell's time, a presumption of political absolutism still makes frequent appearances in theoretical formulations, with the treatment of public debt as sovereign debt being but one illustration among many.

It is sensible to speak of an autocrat as being indebted, just as it is sensible to speak of a person or a corporation as being indebted. For a person or organization to be indebted, there must be some other person or organization to whom the debt is owed. Under these circumstances, it is reasonable for a potential creditor to be concerned with the ability of a potential borrower to repay a loan. It is plausible that someone might become so heavily indebted as to make payment impossible and default unavoidable. Such thinking about the burden of debt servicing is commonly extended to governments, and is expressed in concerns about whether public debt can become so high as to render default likely (see, for instance, Spaventa (1988)). These concerns are often expressed in terms of public debt as a share of GDP, and with the degree of concern escalating with increases in the ratio of public debt to GDP.

Such concerns are generally misleading, however, when it comes to democratic debt, particularly when that debt is held internally. Public debt in a democracy is distinct from public debt in an autocracy. An autocrat must borrow from an outsider; autocratic debt must be held externally. While democratic debt can be held externally, a good deal of it is held internally. The limiting case is where all public debt is held internally. In this case the state becomes a form of financial intermediary that organizes and maintains a complex transaction among the citizenry. The state itself is not indebted, but rather it simply manages the debtor-creditor relationships among the citizenry that arise out of the state's budgetary process. A failure of a state to service public debt is a failure of intermediation. It is the same as with a bank. A depositor might not be able to withdraw his deposit because the bank's loans did not perform as well as the bank expected. Financial intermediation, whether organized privately or publicly, might not work fully in congruence with the expectations of those who participate in that intermediation. These concerns about intermediation, however, are not concerns about the volume of intermediation in relation to some measure of aggregate output. Rather they are concerns that are addressed by such matters as the theory of agency.

Suppose an increase in state spending is financed by debt rather than by increased taxation. Had taxation been selected, there would doubtless exist some taxpayers who would prefer to borrow to cover their added tax payments. The more widespread this preference, the denser would be the resulting network of market-based credit transactions. Public debt replaces this private network of credit transactions with a state-organized program of financial intermediation. The state serves as an intermediary between that part of the citizenry who are borrowers and that part who are lenders. This resulting substitution of state intermediation for market-based intermediation may be generally beneficial, or it might be beneficial to some and harmful to others. Some of those borrowers might have been willing borrowers while others were forced to do so. The extent of voluntary and forced borrowing will depend on a wide variety of rules and institutions that constitute the political and fiscal process.

The fiscal literature contains two polar types of models or frameworks of democratic governance, which may be described as consensual and factional. The most prominent formulation of consensual democracy in the fiscal literature is Knut Wicksell's (1896) formulation of just taxation, along with the associated literature on the benefit principle of public finance. Within Wicksell's formulation, fiscal choices would be made within a legislature that was selected through proportional representation, with the legislature bound by a voting rule of near-unanimity. A further feature of Wicksell's framework is that proposals to spend would be considered simultaneously with proposals to tax, under a type of generalized earmarking.

There are many models of factional democracy in the fiscal literature. They all involve formulations grounded in a process of subordination and domination, whereby some people gain at the expense of others. One such set of models operates with simple notions of majority voting. The archtypical model of majoritarian democracy is one where a winning majority approves larger spending programs because they are able to impose a good share of the cost on the remainder of the citizenry, who do not value the project as highly. Other models of democracy entail processes by which a well-organized and relatively intense minority is able to dominate a poorly-organized and relatively passive majority, as illustrated by the literature on rent seeking (see, for instance, Mitchell and Simmons (1994) and Tullock (1967)).

The analytics of democratic debt differ in important respects, depending on whether a model of consensual or factional democracy is more appropriate. In either case, though, democratic debt differs from sovereign debt, in that it is misleading to speak of the state as being indebted. In a democracy a state can be neither a debtor nor a creditor. The state is rather an intermediary that brings together different parts of the citizenry, though the intermediary may act consensually or factkmally depending on the institutional framework.

4. CONSENSUAL DEMOCRATIC DEBT

The Wicksellian institutional framework is one illustration of a framework whereby fiscal outcomes would reflect generally an underlying consensus among the members of the polity (for further elaboration see Backhaus (1992)). It can be asked under what circumstances in such a polity its members would support debt finance over tax finance. For public debt to arise in a consensually democratic setting, there must exist circumstances under which people would prefer to organize debtor-creditor relationships collectively rather than through market processes. Public debt in this instance would represent a nationalization of financial intermediation, or at least that part of intermediation that arises in response to the claims of extraordinary public finance.

Are there circumstances under which it is plausible that there could exist gains from trade from the collective organization of financial intermediation? A necessary condition for this to occur is that collective intermediation is a lower cost alternative to market-based intermediation. This lower cost provides the potential gains from trade through collective intermediation. Whether that potential would actually be realized in practice is a different matter. Nonetheless, the potential for gain requires some cost advantage for collective intermediation. To the extent there is some element of fixed cost that is independent of the size of a transaction, a collective loan potentially can exploit some economy of scale, as de Viti de Marco argues (1936, pp. 377-398).

In this respect, it is often noted that public debt carries a lower interest rate than private debt. Public debt allows people to borrow at the government's

borrowing rate, which is lower than what people could obtain through market transactions. The question, though, is whether this difference between public and private borrowing rates truly represents a cost advantage for collective intermediation. Various considerations from theories of economic organization and bureaucracy present reasons for being skeptical about this possible cost advantage. Public loans might carry a lower interest rate than private loans even if collective intermediation is more costly. The interest rate on private loans must include the cost of attracting capital into financial intermediation, in competition with other uses of capital. For public loans, or for government generally, there is no explicit cost of capital, even though there is always an opportunity cost of capital. For public loans, the burden of higher cost and greater risk is borne by taxpayers and not by lenders. Unlike the limited liability of corporate shareholders, taxpayers have unlimited liability.

The possibility that the state could serve as an efficient intermediary for the organization of credit transactions in cases of extraordinary public finance cannot be denied. At the same time, however, the interest rate differential between public and private loans cannot be used as evidence in support of the claim that public loans result because the state is an efficient intermediary. Public loans may well result even if the state is an inefficient intermediary, depending on the performance characteristics of various political and fiscal institutions.

To be sure, an argument about economies of scale is not conclusive on this point. Another question concerns whether the liability for subsequent amortization is assigned explicitly at the time the debt is created or is left as a contingency to be determined in subsequent years. In the former case, public debt would operate just as private debt. At the time the debt was created, people would be assigned a schedule of their future liability for amortization. One difficulty with this procedure is that some people may find themselves unable to make their payments in subsequent years. If this happens, the revenues collected will not be sufficient to service the debt. In a corporate setting, the shareholders would be liable for this gap between revenues and expenses. If taxpayers in general are to be liable for this gap when it comes to public loans, a taxpayer's liability becomes contingent on future circumstances and is not genuinely determined at the time the debt is created. For any taxpayer at the time public debt is created, future payments for servicing the debt are contingent on future economic circumstances. It is conceivable that something like Wicksellian near-unanimity could obtain in this setting, but the setting for choice would be more complex than normal credit transactions. A taxpayer's liability for a decision to borrow rather than to tax is not specified currently but is contingent on future circumstances. A certain stream of future payments is replaced by some expectation and associated variance. Under normal presumptions about risk aversion, the greater the variance the greater must be the

cost advantage for public loans, to make public loans a likely outcome in a consensual fiscal setting.

5.   FACTIONAL DEMOCRATIC DEBT

Even if the institutional framework corresponds to some non-Wicksellian framework of factional democracy, as characterized by various models of rent seeking and rent extraction, Ricardian Equivalence must provide a point of departure for any analysis of public debt. Ricardian equivalence must hold in the aggregate simply as a matter of arithmetic. Such an aggregate condition must not be confused with a proposition of behavioral invariance to particular institutional conditions. Aggregative equivalency does not imply that choices between debt and taxation will be invariant to the institutional setting within which such choices are made. The choice between debt and taxation can matter for *particular* people. Indeed, these differences are central for any effort to understand the creation of public debt in the first place. Different institutional settings may lead to different fiscal and budgetary choices, and with different consequences resulting, despite the underlying constraint implied by Ricardian Equivalence.

To illustrate this point, consider the simple model of budgetary equilibrium that is set forth in James Buchanan (1964). Suppose the government provides a single service financed by a proportional income tax, and with public debt precluded through constitutional provision. All citizens have identical preference patterns and the demand for the public service is characterized by unitary price and income elasticities of demand. In this setting, any voting rule gives the same outcome as unanimity; all citizens agree on the size amount of public output and each pays a tax-price equal to his marginal evaluation of public output.

How might the elimination of the constitutional constraint on public borrowing affect the resulting budgetary outcome? For deficit finance to affect budgetary outcomes, public debt would have to be regarded as a lower cost alternative to taxation by some decisive subset of the population, even if collective intermediation does not have a cost advantage over private intermediation. By virtue of the Ricardian theorem, the aggregate present value of future taxes must equal the amount of the budget deficit. But it does not follow that such present-value equivalence holds across individuals, and if it does not hold, the consequences for budgetary choice will depend on the way in which fiscal institutions shape and constrain processes of budgetary choice.

An assumption of full intergenerational altruism has often been used to support Ricardian Equivalence. This assumption converts a model where people have limited lives into one where they live forever and, hence, cannot escape bearing the future consequences of present choices. There is no doubt that

strong intergenerational altruism characterizes some family settings, and models of infinite livelihood probably have great descriptive value in these cases. But there is equally no doubt that intergenerational dislike also characterizes numerous family settings. There are simply too many resources involved in family and divorce law, and in dealing with battery, brutality, incest, and the like to argue otherwise. And there is surely a broad spectrum of in-between cases of varying degrees of benignity, indifference, and the like. It is quite plausible to presume that people differ in the degrees to which they carry intergenerational altruism.[6]

When people differ in their intergenerational altruism, borrowing reduces the relative cost of government services to people the weaker is that altruism. In a simple median voter model, the person whose intergenerational altruism is median within the population will control the budgetary choice. The introduction of a deficit financing option will lead to an expansion in the size of government because it reduces the cost of government to the median voter. A new budgetary equilibrium will be established where, for the median voter, the marginal value of added public output equals his marginal cost through deficit finance. Consider, for instance, three taxpayers of different ages, which can be characterized in terms of taxpaying life expectancy. The eldest taxpayer has 10 years of taxpaying life expectancy, the middle taxpayer has 20 years, and the youngest taxpayer has 40 years. For each of the three, the alternative to a $1,000 tax is to issue public debt in the form of a perpetuity, which requires $100 per year to service. The relevant rate of interest is 10 percent. For the youngest taxpayer, the present value of debt finance is $978, which is practically the same as taxation. For the eldest person, however, the present value of debt finance is only $614. So long as intergenerational altruism is incomplete, public debt becomes systematically less costly with increasing age. Even though negative bequests cannot be left privately, because debts cannot be passed on to heirs, public debt can serve as a means of doing this.

There are other models of collective choice besides the median voter model, and some of these would give descriptively different but analytically similar results. For instance, a ruling political party could be viewed as expanding different tax sources so as to equalize political resistance at the respective revenue margins.[7] The introduction of a debt option lowers marginal political cost. This leads to deficit finance, and the more fully debt is used the higher becomes the political cost of deficit finance. The political pressures from different revenue sources will be equalized at the relevant political margins, where the future taxes represented by debt finance encounters the same political resistance encountered by present taxes. The Ricardian proposition must hold as a condition of political equilibrium, for otherwise there will be a shift in the mix of tax instruments toward those that entail lower political cost.

In any case, public debt becomes a method for transferring wealth among the members of a generation. People with relatively weak bequest motives promote the use of public debt over taxation as a method of increasing their net wealth. People with relatively strong bequest motives suffer a wealth loss through the larger than desired public sector that results. They also recognize that the growth in government debt impinges upon the future well being of their heirs, and so would be predicted to increase their saving in response to the creation of public debt. The creation of public debt does not increase aggregate wealth, but it does increase wealth for some people who are influential at the margins of budgetary choice, while reducing wealth for those who are on the losing side.

Efforts have also been made to explain public debt as a strategic instrument within a model of partisan political competition (see, for instance, Persson and Svensson (1989), Tabellini and Alesina (1990), and Alt and Lowrey (1994)). So long as it is presumed that a government will not repudiate public debt, save for the repudiation of real value that can occur through inflation, an increase in public debt today can restrict the budgetary options of the party in power tomorrow. The party in power may prefer a budget mix that is heavily weighted toward projects of capital construction, while the party out of power may prefer a mix that is heavily weighted toward spending on welfare-like measures. Even if the party in power prefers to operate with a balanced budget, it may enact deficits if it thinks the opposition party has a good chance of coming to power. The budgetary claims to service the debt will impinge upon the ability of the other party to support welfare-like measures. To be sure, the ability of public debt to serve such a strategic purpose depends on the presumption that debt repudiation is not an option. The prospects for repudiation in such a framework would, in turn, seem to depend on the relative concentration of debt holders in one party or the other. If the debt is held largely by supporters of the party in power, debt repudiation by the opposition party could simultaneously expand the options for supporting welfare-like measures and impose a wealth loss on supporters of the other party.

6.   MONEY, SEIGNIORAGE, AND PUBLIC DEBT

Public debt may serve as a form of deferred taxation; however, it may also serve instead as a current tax on money balances. Under contemporary institutions, the monetary base is expanded when a central bank buys public debt. Similarly, the monetary base is reduced by central bank sales of public debt. The effects of transactions in public debt depend on who does the transacting. If debt transactions involve the central bank, monetary policy is being conducted because the monetary base is being changed. If private citizens are doing the transacting, asset portfolios are being changed but the monetary base

is not affected. A government issue of public debt that is purchased by private citizens has no monetary impact. By contrast, an issue of public debt that leads the central bank to increase its holding of public debt increases the supply of money.

Public debt is thus fiscally schizophrenic. Its economic character and impact depends on who is buying or selling it. One day, transactions in public debt may be a means of taxing money balances. On another day, transactions in public debt may be a collective act of financial intermediation that allows a good number of people to defer their income taxes. Which it is, and when, depends on the type of monetary institutions that are in place within a society.[8] The contemporary institutional framework of central banking confounds state borrowing and money creation. Free banking as an alternative to central banking would avoid this confounding of debt and money, by creating a clear separation between state borrowing and money creation. A freezing of the monetary base would also avoid this confounding, at the same time would maintain a central bank. In any case, the fiscal analytics of borrowing and money creation would thus depend on both political regime and monetary framework.

While free banking has appeared in a number of historical instances, central banking is clearly a predominant feature of monetary arrangements. There are two broad approaches one might take to explaining the dominance of central banking. One revolves around claims of market failure. These claims are represented by the aphorism that money can't manage itself. The other approach to explaining the dominance of central banking involves processes of political domination. Money may well be able to manage itself through free banking, as Selgin (1988 argues), but it is not allowed to do so because there are political profits to be reaped through central banking (Wagner 1986b). While a detailed examination of these contending approaches to explanation are outside the scope of this essay, some brief consideration can be given as to how central banking might serve as an instrument of domination and subordination within a framework of factional democracy.

Central banking allows for some expansion in the government's budgetary capacity, as compared with what that capacity would be under free banking. It does this by creating an instrument of taxation that does not exist with free banking. This instrument is the taxation of people's money balances (Friedman (1971), Selgin and White (1999)). To be sure, it is possible to tax money without central banking. It would be relatively costly to do so, however, as is illustrated historically by various practices where coins were debased as they passed through royal treasuries. With central banking and fiat money, it is nearly costless for the treasury to tax money. An expansion in the nominal stock of money reduces the real value of existing units of money and provides revenue to the central bank in the process.

Whether this revenue generated through inflation accrues to the central bank or the treasury depends on various institutional features that govern the relations between the two. In some places the central bank is located within the treasury, in which case the revenues accrue directly to the treasury. In other places, the central bank has some degree of independence from the treasury. This independence may allow the central bank to use seigniorage for its own purposes. Typically, though, most of the revenues are returned to the treasury in what appears to be a voluntary transfer. The degree of central bank independence is a political outcome, and that independence can continue only so long as there is not strong political interest to remove that independence. A "voluntary" return of seigniorage is surely a means of keeping support for nominal independence, and with the central bank securing some seigniorage for its own use. In other words, the central bank can claim some share of seigniorage revenues for its own uses, so long as it raises a satisfactory amount of revenue for the government (Boyes, Mounts, and Sowell (1998) and Toma (1982)).

To be sure, there is some question of the extent to which inflation is pursued directly, as against its being a by-product of the pursuit of other outcomes and the promotion of other interests through government. In the former case, seigniorage is pursued directly as a source of tax revenue. In the latter case, inflation results as an adjustment to other policy measures. Cost-push inflation has little merit on purely economic grounds. An increase in prices in particular sectors cannot be a direct source of inflation, because output will expand and prices will fall elsewhere in the economy. However, cost-push inflation may acquire added explanatory power once political processes and interests are taken into consideration (see, for instance, Iversen (1999)). A simple framework could involve unions and a central bank. Unions place a positive value on increases in real wages among members and a negative value on unemployment among members. The central bank evaluates both inflation and unemployment negatively. Increases in real wages that would otherwise increase unemployment might induce the central bank to increase the monetary base to reduce the rise in unemployment that would otherwise result. In doing this, the central bank is judging this course of events to be preferable by its own calculus to the course of events that would otherwise result. That alternative course would start with rising unemployment in the unionized sectors, which in turn would induce shifts of labor elsewhere, as well as inspire entrepreneurial efforts to organize lower-priced substitutes for union-produced products.

While central banks are normally associated with changes in the stock of money, they are also heavily implicated in processes of credit allocation. The pattern of credit allocations includes a significant variety of governmental regulations that influence the allocation of credit. In modern democratic states where interest group political processes are woven throughout the economy,

credit allocation would seem clearly to offer more scope for a market for legislation to operate than can monetary policy. After all, the stock of money-based credit is on the order of 10 to 15 times larger than the stock of money. Moreover, credit involves not even a hint of neutrality, but is wholly concerned with individual wealth positions. Governments are involved in numerous ways in altering market-based credit terms and relationships. Usury regulations have, of course, been around for a long time, and more recently a variety of "fairness" regulations have sought to increase the supply of credit to certain racial and income categories.

Consider an effective political desire to increase the credit supplied to some particular interest group. One way this shift in credit can be accommodated is for a reallocation of credit away from other participants. This policy measure would be a regulation-imposed form of tax-transfer operation, with the regulatory agency forcing lenders to expand their lending to favored groups, and financing that lending by reducing their lending to unfavored groups. This type of operation is clearly possible, as tax and transfer operations are at the core of interest group processes of political competition.

There is, however, an alternative possibility. The credit expansion to the favored group can be accommodated by the central bank through an expansion in the stock of high-powered money. Such a monetary expansion would reduce the political opposition to the credit reallocation that would otherwise result in the absence of an accommodating monetary policy, principally because the cost of the credit reallocation is shifted away from other interest groups onto the population generally. Monetary policy thus becomes the equilibrating vehicle that accommodates changes in the market for credit that emanates from interest group politics, similar to short-term capital movements serving to create equilibrium in a balance of payments. Credit might drive money within an interest group model of government, because that is where the greatest harvest of political profit lies, even if it is money that drives credit in a liberal market economy.

7.   CONCLUDING REMARKS

The economic analysis of public debt differs depending on the presumed political setting. It may be reasonable to characterize public debt choices within an authoritarian regime as being made by a single mind, but such a characterization is surely inapt for democratic regimes. To be sure, even in such regimes Ricardian equivalence must hold in the aggregate, but this aggregate equivalence is irrelevant for human conduct in fiscal choice. Deficit finance injects a systematic differential among current citizens in the cost of public finance, making that cost lower the weaker the degree of intergenerational altruism and the higher the rate of time preference.

In this chapter I have considered public debt as an alternative to tax finance. But within existing monetary institutions, public debt creation often serves as a disguised form of money creation. The possibility of inflationary finance opens up, in turn, new avenues along which deficit finance may serve as a means by which politically dominant groups are able to impose costs on others. A complete analysis of public debt within an interest-group approach to fiscal processes will clearly have to incorporate and integrate such monetary considerations, at least under prevailing central banking institutions.[9] Wherever such an analysis might lead, Ricardian equivalence will have to hold in the aggregate; yet such aggregative equivalence will be only a side show in the fiscal drama that public debt represents.

NOTES

1. It can also use regulation as a nonfiscal substitute for what would otherwise be a budgetary operation. For instance, a state could pretty much abolish its education budget by requiring parents to send their children to designated schools.
2. On Cameralist budgetary practice, see Backhaus and Wagner (1987). For a comparison of Justi and Smith, see Wagner (Forthcoming).
3. For a careful statement of this general theme, see Buchanan (1967).
4. For a sampling of analyses in this vein, see Alesina and Perotti (1994), Buchanan and Roback (1987), Congleton (1992), Cukierman and Meltzer (1989), and Tabellini and Alesina (1990).
5. The importance of institutional settings for debt analysis is explored in Wagner (1986a, 1996).
6. In a related line of argument, Cukierman and Meltzer (1989) model public debt as a means by which people who wish to leave negative bequests can do so, despite the prohibition against doing so privately. In their framework, debt choices are driven by intergenerational wealth redistribution, whereas here debt choices emerge out of contemporary budgetary politics.
7. Such an approach to tax politics is sketched in Hettich and Winer (1999).
8. For a valuable analysis of monetary institutions that has considerable relevance for the analytics of public debt explored here, see Lawrence White (1999).
9. For a small sample of work in this area, see Grier and Nieman (1987), Tabellini (1987), and Parkin (1986).

REFERENCES

Alesina, A. and Perotti, R. (1994). "The Political Economy of Budget Deficits." *IMF Staff Papers* 42: 1-37.

Alt, J. and Lowrey, J. (1994). "Divided Government, Fiscal Institutions, and Budget Deficits: Evidence from the States." *American Political Science Review* 88: 811-828.

Backhaus, J. G. (1992). "The State as a Club: A Perspective for Public Finance in a Prosperous Democracy." *Journal of Public Finance and Public Choice* 10: 3-16.

Backhaus, J. G. and Wagner, R. E. (1987). "The Cameralists: A Public Choice Perspective." *Public Choice* 53: 3-20.

Barro, R. J. (1974). "Are Government Bonds Net Wealth?" *Journal of Political Economy* 1095-1118.

Barro, R. J. (1979). "On the Determination of the Public Debt." *Journal of Political Economy* 87: 940-971.

Boyes, W. J., Mounts, W. S., and Sowell, C. (1998). "Monetary and Fiscal Constitutions and the Bureaucratic Behavior of the Federal Reserve." *Public Finance Review* 26: 548-564.

Brennan, G. and Buchanan, J. M. (1980). *The Power to Tax: Analytical Foundations of a Fiscal Constitution.* Cambridge: Cambridge University Press.

Buchanan, J. M. (1958). *Public Principles of Public Debt.* Homewood, IL: Richard D. Irwin.

Buchanan, J. M. (1967). *Public Finance in Democratic Process.* Chapel Hill: University of North Carolina Press.

Buchanan, J. M. (1964). "Fiscal Institutions and Efficiency in Collective Outlay." *American Economic Review,* Papers and Proceedings 54: 227-235.

Buchanan, J. M. and Roback, J. (1987). "The Incidence and Effects of Public Debt in the Absence of Fiscal Illusion." *Public Finance Quarterly* 15: 5-25.

Bulow, J. and Rogoff, K. (1989). "Sovereign Debt: Is To Forgive To Forget?" *American Economic Review* 79: 43-50.

Calvo, G. (1988). "Servicing Public Debt: the Role of Expectations." *American Economic Review* 78: 647-661.

Congleton, R. (1992). "The Politics of Debt." *Journal of Public Finance and Public Choice* 10: 17-34.

Cukierman, A. and Meltzer, A. H. (1989). "A Political Theory of Government Debt and Deficits in a Neo-Ricardian Framework." *American Economic Review* 79: 13-32.

de Viti de Marco, A. (1936). *First Principles of Public Finance.* New York: Harcourt Brace.

Drazen A. (2000). *Political Economy in Macroeconomics.* Princeton, NJ: Princeton University Press.

Eaton, J. and Gersovitz, M. (1981). "Debt with Potential Repudiation: Theoretical and Empirical Analysis." *Review of Economic Studies* 48: 289-309.

Ferguson, J. M. (ed.) (1964). *Public Debt and Future Generations.* Chapel Hill: University of North Carolina Press.

Friedman, M. (1971). "Government Revenue from Inflation." *Journal of Political Economy* 79: 846-856.

Grier, K. B. and Nieman, H. E. (1987). "Deficits, Politics, and Money Growth." *Economic Inquiry* 25: 201-214.

Grossman, H. I. and Van Huyck, J. B. (1988). "Sovereign Debt as a Contingent Claim: Excusable Default, Repudiation, and Reputation." *American Economic Review* 78: 1088-1097.

Hettich, W. and Winer, S. L. (1999). *Democratic Choice and Taxation.* Cambridge: Cambridge University Press.

Horrigan, B. R. (1986). "The Determinants of the Public Debt of the United States, 1953-1978." *Economic Inquiry* 24: 11-23.

Iversen, T. (1999). "The Political Economy of Inflation: Bargaining Structure or Central Bank Independence?" *Public Choice* 99: 237-258.

Justi, J. H. G. von. (1969). *Natur und Wesen der Staaten.* Darmstadt: Scientia Verlag Aalen. [Reprint from 1771 edition.]

Lerner, A. P. (1948). "The Burden of the National Debt." In *Income, Employment, and Public Policy,* Metzler, L. A., ed., pp. 255-275. New York: W. W. Norton.

Mitchell, W. C. and Simmons, R. T. (1994). *Beyond Politics: Markets, Welfare, and the Failure of Bureaucracy.* Boulder, CO: Westview Press.

Parkin, M. (1986). "Domestic Monetary Institutions and Deficits." In *Deficits,* Buchanan, J. M., Rowley, C. K., and Tollison, R. D., eds., pp. 310-337. Oxford: Blackwell.

Persson, T. and Svensson, L. E. O. (1989). "Why a Stubborn Conservative Would Run a Deficit: Policy with Time-Inconsistent Preferences." *Quarterly Journal of Economics* 104: 325-345.

Ricardo, D. (1817) [1963]. *Principles of Political Economy and Taxation.* Homewood, IL: Richard D. Irwin.

Schumpeter, J. A. (1954). *A History of Economic Analysis.* New York: Oxford University Press.

Selgin, G. (1988). *A Theory of Free Banking.* Totowa, NJ: Rowman and Littlefield.

Selgin, G. and White, L. H. (1999). "A Fiscal Theory of Government's Role in Money." *Economic Inquiry* 37: 154-165.

Smith, A. (1776) [1937]. *An Inquiry into the Nature and Causes of the Wealth of Nations.* New York: Modern Library.

Spaventa, L. (1988). "Introduction: Is There a Public Debt Problem in Italy?" In *High Public Debt: The Italian Experience,* Giavazzi, F. and Spaventa, L., eds., pp. 1-24. Cambridge: Cambridge University Press.

Tabellini, G. (1987). "Central Bank Reputation and the Monetization of Deficits." *Economic Inquiry* 25: 185-200.

Tabellini, G. and Alesina, A. (1990). "Voting on the Budget Deficit." *American Economic Review* 80: 37-49.

Toma, M. (1982). "Inflationary Bias of the Federal Reserve System." *Journal of Monetary Economics* 10: 163-190.

Thompson, E. A. (1967). "Debt Instruments in Both Macroeconomic and Capital Theory." *American Economic Review* 57: 1196-1210.

Tullock, G. (1967). "The Welfare Economics of Tariffs, Monopolies, and Theft." *Western Economic Journal* 5: 224-232.

Wagner, R. E. (1986a). "Liability Rules, Fiscal Institutions, and the Debt." In *Deficits,* Buchanan, J. M., Rowley, C. K., and Tollison, R. D., eds., pp. 199-217. Oxford: Basil Blackwell.

Wagner, R. E. (1986b). "Central Banking and the Fed: A Public Choice Perspective." *Cato Journal* 6: 519-538.

Wagner, R. E. (1996). "Who Owes What, and to Whom? Public Debt, Ricardian Equivalence, and Governmental Form." *Review of Austrian Economics* 9: 143-157.

Wagner, R. E. (Forthcoming). "The Cameralists: Fertile Sources for a New Science of Public Finance." In *Founders of Modern Economy: Masstricht Lectures in Political Economy,* Backhaus, J. G., ed. Cheltenham: Edward Elgar.

White, L. H. (1999). *The Theory of Monetary Institutions.* Oxford: Blackwell.

Wicksell, K. (1896). "A New Principle of Just Taxation." In *Classics in the Theory of Public Finance,* Musgrave, R. A. and Peacock, A. T., ed., pp. 72-118. London: Macmillan, 1958.

Chapter 9

REGULATION BY TAXATION*

Andy H. Barnett
Auburn University, American University of Sharjah
abarnett@business.auburn.edu

Bruce Yandle
Clemson University, Political Economy Research Center (PERC)
yandle@clemson.edu

Abstract	This chapter examines some cases where taxation seems perhaps to be supported more as a means of accomplishing specific regulatory ends than as a means of raising general revenues. Many of these instances involve claims about the control of external costs, as through various environmental taxes and fees.
Keywords:	Alcoholic beverage taxation, Pigouvian doctrine, regulatory taxation, tax relief, taxing environmental use
JEL classification:	H20, L50

1. INTRODUCTION

In 1971, Richard A. Posner (1971) wrote "Taxation by Regulation," a paper that sought to explain some aspects of regulation by considering theories of taxation."[1] In a reversal of the Posner paper, we seek to explain the occurrence of discriminatory excise or "sin" taxes by drawing on theories of regulation. We note at the outset that (1) all regulations tax those impacted by the regulations and (2) all taxes discourage the taxed activity and, therefore, have regulatory impacts. From the standpoint of the affected individual or firm, all regulations are taxes and all taxes regulate. Since that is the case, it is public-purse revenue generation that distinguishes a regulatory tax from a regulation

---

that might bring the same behavioral outcome. A government may choose either to write and enforce detailed rules affecting the consumption of alcoholic beverages or to impose well-modulated taxes that lead to the same level of consumption. The difference between the two regulatory regimes is found in the prospects for revenues net of enforcement and collection costs.

We call attention to the fact that the use of excise or regulatory taxes goes back centuries (Shughart, 1997, Sirico, 1995), but their initial economic analysis is primarily associated with the work of A. C. Pigou (1938).[2] It was Pigou who focused on the political prospect of using taxes (and subsidies) for the expressed purpose of adjusting market outcomes judged to be less efficient or politically desirable than what might result from a purely competitive market. Indeed, Pigou's name is so prominently associated with regulation by taxation that today those who support any form of market intervention by government are often called Pigouvian. As we shall see, Pigou himself may not have been as "Pigouvian" as some of his modern disciples. Indeed, he is perhaps more remarkable for explaining why his taxation ideas will not work than for suggesting their use in the first place.

In this chapter, we first discuss theories of taxation as regulation. As the discussion unfolds, we arrive at a point where the forces of rent seeking and special interest demand for regulation subvert a public interest-based regulatory taxation process. What starts as an efficiency story ends as a special interest regulation story. Our taxation theory relies heavily on the regulatory counterpart. The next section presents Pigou's treatment of the topic. Making heavy use of direct quotations from Pigou, we show that he clearly understood the political economy of rent seeking and ultimately disavowed the possibility of successfully applying his own Pigouvian taxes in a real world context.

Pigou's disavowal of Pigouvian taxes has certainly not affected the apparent use of excise taxes for regulatory purposes, found notably in conjunction with alcoholic beverages and environmental emissions. Pigou refers to alcoholic beverages and air pollution in building the theoretical case for the use of corrective taxes while at the same time pointing out how the politics involved will distort their purpose. This introduces the chapter's next-to-last section, which is an empirical examination of alcoholic beverage taxation across the 50 U.S. states and an investigation of environmental taxes among 21 OECD countries. We close the chapter with some final thoughts on taxation as regulation.

2.   A THEORY OF REGULATORY TAXATION

A liberal government with citizen-delegated authority has three instruments available to achieve its goals. One device is force; the government can regulate. Another is government production; it can deliver justice, build and operate highways, and provide water and sewer systems. A third is taxes and subsidies;

government can use its fiscal powers to impose excise taxes on things deemed undesirable and provide subsidies for things meritorious. Just who establishes priorities and how fiscal targets are identified is an important matter that we leave for another time.

Obviously, all three instruments when purposefully designed and applied can lead to the achievement of a desired goal, at least theoretically. That is, in theory the instruments can be used to serve a broad public interest. However, the distributional consequences of the three mechanisms are likely to be quite different. In addition, the ease with which bureaucrats, legislators, and citizens can pursue political favors that subvert the public weal may differ substantially across the three tools.

As a starting point, we note that all government production must involve some mechanism to acquire resources and, hence, cannot be considered apart from the tax required to finance the production. Government revenues can arise from public ownership and sale of natural resources such as timber, oil, and gas. However, government ownership prevents private rights from evolving. This limitation on the evolution of property rights has social cost and can generate impacts that are qualitatively equivalent to taxes or regulations. In short, government production necessarily requires some form of taxation. Opportunity cost will be paid one way or another.

To focus our inquiry, consider the relative merits of regulation and taxation. Assume for the moment that the administrative costs for managing a regulatory regime for some stated purpose are the same for either instrument. Assume further that the burden of the tax borne by consumers is the same as the burden of the regulation that accomplishes the same end. In other words, all costs are the same for the two instruments.

Taxation gives politicians revenues and some measure of control over the taxed commodity. Loopholes and special tax treatment can be developed, modified easily and auctioned or withheld politically to reward and punish particular groups (Stigler, 1971, McChesney, 1987). Then, with lobbying revenues pocketed by the politicians, the government revenue collected from the tax can be used to gain further control, perhaps over other commodities. Regulation and regulation modification, on the other hand, can also be auctioned to special interest groups but gives only control. But there is no stream of tax revenues that follows a pure regulatory auction. From the regulator's perspective, a regulatory tax provides a "double dividend."[3]

Continuing with the assumption of neutrality with respect to administrative costs and the political advantage we claim for taxation as regulation, we must ask what characteristics will define the targeted activity taxed and thereby regulated? Why are some activities especially suited for regulation by taxation while others are better suited for direct regulation or public production? To be more specific, why are tobacco products, alcoholic beverages, and gasoline

taxed frequently for acclaimed regulatory purposes? If a quantitative or qualitative outcome is the purpose of the tax, why not simply regulate or have the state produce and market the products? A major part of the answer to these questions must relate to the net revenue potential associated with taxation.

Generally speaking, it seems that excise or "sin" taxes systematically fall on commodities for which demand is highly price inelastic. Efficiency considerations suggest this is beneficial if revenue enhancement is the purpose of the tax. As first analyzed by Ramsey (1927) and discussed in the context of sin taxes by Grossman, Sindelar, Mullahy, et al. (1993) and Shughart (1997), price inelastic demand curve yields a smaller tax burden when excise taxes result in higher prices. A tax authority seeking to minimize the burden of taxation would levy taxes inversely with respect to price elasticity. The most inelastically demanded goods would carry the largest share of the taxes. Of course, this same logic contains a paradox. If the purpose of the tax is to reduce an undesirable activity, then the more price inelastic the demand the less successful the regulatory tax. If diminution in the regulated activity is the goal, regulation is preferred to taxation of inelastically demanded goods. Discouraging consumption could be the primary objective of such taxes only for goods where demand is relatively price elastic. Since most goods on which excise taxes have been levied (tobacco, alcohol, gasoline, and luxury items) are goods for which demand is relatively price inelastic, one might then draw the conclusion that the impact (and perhaps the intent) of most excise taxes is revenue enhancement, not a meaningful diminution in the activity being taxed.[4]

Of course, how things end up does not necessarily explain how they started. At some point in history, politicians had to learn about the relative productivity of various taxes. No one knew about elasticities. The revenue effects of different tax schemes had to be discovered. This suggests a political sorting took place. In a theoretical sense, the need for revenue called for taxes that were placed randomly on different commodities. Experimentation generated knowledge about which commodities or activities were the most productive revenue producers. Sorting took place, and explanations were offered for keeping certain taxes and eliminating others.

The theory we have just sketched is supported partly by the story told by Brenda Yelvington (1997) in her account of the 18th century Whiskey Rebellion and other early American experience with sin taxes. Her summary of the rebellion over whiskey taxes tells us that:

[T]he events surrounding the first sin tax are important for several reasons. First, the new government demonstrated its ability to enforce the law. Second, taxpayers displayed their willingness to fight what they believed were unjust taxes. Third, and possibly most important, Congress discovered a new source of revenue—the selective excise tax. It had been proved that taxes on articles of consumption could be successfully imposed if the articles themselves were

portrayed as luxuries or if their use were pronounced harmful to individuals or the economy as a whole (p. 34).

But her later discussion of Civil War taxation is more to the point. As a result of the requirements to fund the war, the federal government placed excise taxes on all manufactured items, legal documents, financial transactions, transportation system receipts, advertisers. Then, Yelvington (1997, p. 37) reports:

> "As the federal government's revenue needs continued to grow, Congress demonstrated its willingness to raise excise taxes sharply. In the Internal Revenue Act of 30 June 1864, tax rates on distilled spirits were raised from twenty cents to $1.50 per gallon. Subsequent increases in the alcohol tax resulted in 1865 in a top rate of $2.00 per gallon, which was ten times the original cost of the product. Other increases included the tax on loose tobacco, which more than doubled, and the tax on cigars, whose maximum rate jumped from $3.50 per thousand to $40.00 per thousand. As had been done following the War of 1812, the majority of the Civil War excise taxes were rescinded by Congress, with legislation passed in 1867 and 1870. However, the liquor and tobacco taxes remained in place and became permanent fixtures of the federal revenue system … No longer were such taxes justified solely as ways of protecting the country's health and morals. Rather, excise taxes were rationalized as patriotic ways of raising additional revenues during wartime emergencies."

Gifford's account of the early British experience with liquor taxation parallels Yelvington's, but with a slightly different twist. In the British case, the first legislative effort claimed to be about reducing consumption. Gifford (1997, p. 61) describes the evolution of the tax this way:

"In 1736, the British Parliament passed the Act Against Spirituous Liquors, covering all strong spirits. It imposed prohibitive taxes designed to curtail all consumption, but this law also failed, because it was impossible to enforce, was widely ignored, and was the cause of large-scale social disruption and protest that led to its replacement with moderate taxes in 1743. In 1778, the predominant purpose of taxes on spirits shifted to revenue raising—and high taxes to finance the fight against American independence led to widespread illicit distillation and smuggling."

Perhaps the relative resistance to liquor taxation served as a signal that demand curves were inelastic. Without knowing anything about price elasticity, revenue-hungry politicians could experiment and find productive sources for revenue enhancement.

2.1.   Tax Relief as Regulation

Typically when thinking of taxes as regulatory mechanisms we think of levies on undesirable behavior, i.e., Pigouvian taxes. Taxes on pollution can discourage polluters and excise taxes on alcohol, tobacco or gambling can discourage their consumption. But there is another side to the taxes as regulation story. Tax relief can also be used to regulate.

A well-known feature of the American political scene is the propensity of candidates to "buy" votes and interest group support by offering wealth transfers. In his presidential campaign literature, Al Gore (www.algore2000.com. July 24, 2000) offered a wide variety of tax relief measures that target a wide variety of interest groups. Clearly much of the intent of the Gore campaign was the time-honored purchase of votes by the offer of political favors.

On the other hand, it is equally clear that the Gore campaign proposed tax relief to promote regulatory objectives. For example, the Gore campaign web site included the following energy policy offerings:

(1) A tax credit of up to $6,000 for the purchase of more fuel-efficient cars and sport utility vehicles, $5,000 for fuel efficient pick-ups and up to $15,000 for heavy trucks,
(2) Tax credits of $2,000 to homeowners to purchase energy-efficient new homes or $1,000 to retrofit existing homes,
(3) A tax credit to reimburse consumers for purchasing cleaner energy,
(4) A 20 percent tax credit to business for the purchase of energy-efficient building equipment, and.
(5) Tax relief for farmers who produce materials for bio-based fuels.

A scan of the George W. Bush presidential campaign web site revealed a shorter list of specialized tax-relief promises, but emphasized "broad tax cuts," explaining that:

"Every family faces different challenges: some need better childcare, some need tutoring for their children, and others need more after-school programs. Government cannot tailor its programs to the needs of each family. The best way to help all families is to let each family keep more its income and spend it as it deems appropriate."

(www.georgewbush.com/issues.asp?FormMode=fullText&ID=3 (Aug. 1, 2000)

The Bush short list of specialized tax relief included: (1) doubling the child credit, (2) reducing the marriage penalty by increasing the maximum standard deduction for couples, (3) increasing the maximum contribution to education savings accounts from $500 to $5000, and (4) facilitating the deductibility of charitable contributions.

That tax relief is used to regulate should not be surprising. For many years environmental economists pointed to the similarities between incentives created by taxes on pollution and subsidies for pollution abatement. The opportunity cost for targeted decision makers is the same. However, a pollution tax creates resistance from those taxed but garners support from those who seek to punish environmental "sinners" and those who seek additional government revenues. Subsides for abatement, on the other hand, may offer the prospects

of similar support from environmentalists with little or no resistance from polluters.

2.2.   Taxation in the Theory of Regulation

Our discussion of the double dividend provided by regulatory taxes was couched in the language of regulation theory. Politicians were described as auctioning taxes and loopholes to the highest political bidder and then having the resulting tax revenue to use in providing additional special interest benefits. Clearly, the evolved theory of regulation contains a number of positive components that may be related to taxation as regulation. To provide meaningful content to this statement, we must consider the elements that form the theory.

Public interest theory is the first element. Primarily based on a search for efficiency, the public interest theory asserts that when involved in political decisions, politicians shove aside their personal interests and are motivated primarily by a desire to maximize social welfare. Regulatory taxes will be used to correct market imperfections that result from monopoly power, incomplete information, the absence of property rights, or the inability of people to transact. Of course, designing efficient regulatory taxes requires detailed information about the demand and supply of the commodity or activity to be regulated. More often than not, special interest groups are pleased, if not anxious, to deliver the information to the politician. In the process, the human tendencies of politicians can override their primary public interest motivations. Capture theory steps in when public interest theory breaks down. This theory indicates that publicly interested politicians unwittingly are captured by the interests they serve. In effect, the regulated become the regulators.

The special interest or economic theory of regulation (Stigler, 1971, Peltzman, 1976) argues that things are not simple as suggested by public interest and capture. Instead, the market is the more appropriate model. There is demand and supply of regulation. Those interest groups with the most to lose or gain will struggle hardest to obtain favors in the political market place. Politicians are brokers in a process that rewards and punishes the stronger and weaker interest groups. At the same time, no single group can have it all its way. There is always a public interest component that carries weight in the struggle.

The Bootlegger and Baptist theory of regulation (Yandle, 1983, 1999) argues that the more durable regulation emerges when two apparently opposing interests, as in the title, support the same political objective. Baptists who seek a diminution in the consumption of alcoholic beverages support laws that close the retail outlets on Sunday. Bootleggers support the same legislation. With a bit of imagination, canny politicians and interest group members can evoke a public interest smokescreen that provides moral justification of government action that regulates and provides special interest benefits to some of

the regulated. Finally, McChesney's (1987) theory of rent extraction predicts that politicians do in fact obtain contributions by way of threats of action, which are not necessarily carried out. In all of this, the notion of rent-seeking behavior (Buchanan, Tullock, and Tollison, 1980), efforts by individuals and groups to obtain political favors, plays a major role.

These theories of regulation suggest that politicians designing regulatory taxes will discriminate across interest groups and members within those groups to generate differential effects that favor the political powerful. Market share can be adjusted by altering the relative tax rates on competing products. Indeed, some product markets may be almost eliminated by means of discriminatory taxes. The fact that sin taxes fall heaviest on inelastically demanded goods provides an opportunity for politicians to engage bootleggers and Baptists in the pursuit of revenues. Higher taxes can be imposed on gasoline and other fuels in the name of the environment than would be the case otherwise. Finally, once imposed, regulatory taxes can be modified to alter their relative effects or threats of action can be announced, all the while generating the double dividend of revenue and control.

2.3.   Summary

Our theory of regulatory taxation rests heavily on the double dividend generated for politicians who obtain specialized control of the item being regulated and receive tax revenues that can be used to further their political advantage. Inelastically demanded goods and services provide the most fertile ground for imposing regulatory taxes that serve the dual purpose of reducing behavior deemed undesirable while at the same time generating substantial tax revenues.

Among the positive theories of regulation, Bootleggers and Baptists combine with the special interest theory to provide an explanation of the occurrence of regulatory taxation. This composite theory suggests that revenue hungry politicians will seek the help of morally upright citizens in identifying activities that hold great promise for regulatory taxation. For the result to be durable, both parts of the coalition must remain in place. For its survival, there must be bootleggers and "Baptists" who support the regulatory tax.

3.   THE PIGOUVIAN DOCTRINE

How much of this did A. C. Pigou understand? Was Pigou oblivious to political realities? Ask anyone exposed to graduate education in economics to describe the contributions of A. C. Pigou and the response will surely include corrective or Pigouvian taxes. Pigou's taxation idea seems logical enough: When the market process fails to adequately link benefits to costs, politicians, acting through government, can sense the problem, seize the moment, and

impose corrective taxes or subsidies and thereby achieve an increase in so-cial welfare. Described almost religiously in most principles of economics textbooks, the Pigouvian proposal seems perfectly suited for a world full of unwanted external costs and market inefficiencies. This is Pigou strictly inter-preted.

Those who search for Pigou's own words on the matter generally turn to his discussion of marginal social costs and marginal social benefits and read how he related these welfare concepts to their private counterpart. Pigou warns that under certain circumstances, marginal private costs will diverge from marginal social costs. When this occurs, Pigou (1938, p. 192) notes that:

> It is however possible for the State, if it so chooses, to remove the divergence in any field by "extraordinary encouragements" or "extraordinary restraints" upon investments in that field. The most obvious forms which these encouragements or restraints may assume are, of course, those of bounties and taxes.

Pigou then offers alcoholic beverages as an example of what he is talking about:

> The private net product of any unit of investment is unduly large relatively to the social net product in the businesses of producing and distributing alcoholic drinks. Consequently, in nearly all countries, special taxes are placed upon these businesses (p. 192).

Since countries tax the business of alcohol, then the consumption of alcoholic beverages must generate unwanted social costs. Or could there not be another reason for the taxes he observes? In any case, Pigou makes certain that the reader is not trapped into thinking that Pigouvian incentives are just about alcoholic beverages. He extends his model to include automobiles and insur-ance:

> The principle is susceptible of general application. It is employed, though in a very in-complete and partial manner, in the British levy of a petrol duty and a motor-car license tax upon the user of motor cars, the proceeds of which are devoted to the service of the roads. It is employed again in an ingenious way in the National Insurance Act. When the sickness rate in any district is exceptionally high, provision is made for throwing the consequent ab-normal expenses upon employers, local authorities or water companies, if the high rate can be shown to be due to neglect or carelessness on the part of any of these bodies (p. 193).

3.1.   Was Pigou Mugged by Reality?

In later writing, Pigou (1960, p. 99) more knowingly, it seems, repeats his discussion of divergences between marginal social and marginal private cost. He recognizes that the strict case is from a world of zero transaction and rent-seeking costs. He warns:

> When maladjustments have come about or are threatening to come about …, it is always possible, on the assumption that no administrative costs are involved, to correct them by imposing appropriate rates of tax on resources employed in uses that tend to be pushed too

> far and employing the proceeds to provide bounties, at appropriate rates, on uses of the opposite class ….

Of course, in real life considerable administrative costs would be incurred in operating schemes of this kind. These might prove so large as to outweigh the benefit even of the optimum scheme, and a fortiori, of the others. Again, it must be clearly understood that, unless the rate of taxes and bounties imposed fall within certain determined limits, more harm than good will be done even though there are no administrative costs.

Pigou (1938, p. 331) also clearly recognized the possibilities that rent-seeking interests and politicians could destroy any hope of an efficient outcome from a "Pigouvian" tax. Pigou (1938, p. 331) approvingly quotes Bemis: "Every public official is a potential opportunity for some form of self-interest arrayed against the common interest."

He describes the implications of this with considerable care, noting the recognition of what later would be termed "uninternalized external costs" was not sufficient reason to call for government action. The cost of government failure had to be balanced against the costs of market failure.

> In any industry, where there is reason to be believe that the free play of self-interest will cause an amount of resources to be invested different from the amount that is required in the best interest of the national dividend, there is a prima facie case for public intervention. The case, however, cannot become more than a prima facie one, until we have considered the qualifications, which government agencies maybe expected to possess for intervening advantageously. It is not sufficient to contrast the imperfect adjustments of unfettered private enterprise with the best adjustment that economists in their studies can imagine. For we cannot expect that any public authority will attain, or will even whole-heartedly seek, that ideal. Such authorities are liable alike to ignorance, to sectional pressure and to personal corruption by private interest. A loud-voiced part of their constituents, if organized for votes, may easily outweigh the whole. This objection to public intervention in industry applies both to intervention through control of private companies and to intervention through direct public operation. On the one side, companies, particularly when there is continuing regulation, may employ corruption, not only in getting their franchise, but also in the execution of it …. This evil has a cumulative effect; for it checks the entry of upright men into government, and so makes the corrupting influence more free. On the other hand, when public authorities themselves work enterprise, the possibilities for corruption are changed only in form (p. 331).

3.2.   Taxation for Revenue Only

Finally, we note that unlike modern day Pigouvians who uncritically expect government to make things better, Pigou (1960, p. 208) raises a fundamental concern: "[I]t has to be considered whether governments, as constituted in real life, can be trusted, or can trust themselves, with these difficult matters." This remark came in a discussion of the relative merits of assisting infant industries in the larger context of maintaining free trade. Pigou was not about to be

caught in a Pigouvian trap. His own model of regulatory taxes could be used to justify a call for temporary tariffs to protect a fledgling industry on its way to becoming strong enough to compete in world markets. Tossing the whole idea aside, Pigou happily quotes from Sedgewick, and says:

> [B]ut I am, nevertheless, strongly of opinion that it is practically best for a government to adhere to the broad rule of 'taxation for revenue only'—at any rate in a free community where habits of commercial enterprise are fully developed. My ground for this opinion is that I do not think we can reasonably expect our actual governments to be wise and strong enough to keep their protective interference within due limits; owing to the great difficulty and delicacy of the task of constructing a system of import duties with the double aim of raising revenue equitably and protecting native industry usefully, and the pressure that is certain to be put upon the government to extend its application of the principle of protection if it is once introduced. I think, therefore, that the gain protection might bring in particular cases is always likely to be more than counterbalanced by the general bad effects of encouraging producers and traders to look to government aid in industrial crises and dangers, instead of relying on their own foresight, ingenuity and energy; especially since the wisest protection in any one country would tend in various ways to encourage unwise protection elsewhere" (p. 208).

3.3.   Summary

Was Pigou a Pigouvian? We are of the opinion that he was not. While Pigou developed the taxation as regulation theory, and related notable situations for using it, he recognized the difference between public finance theory and the practice of politics. In the end, Pigou's recommendation that taxes be used for revenue alone, not for regulation, is not enough to constrain the politicians who, aided and abetted by special interest groups, will use the strict Pigouvian Doctrine to justify taxes that will indeed provide more revenue to feed the unconstrained appetite for public sector growth.

4.   AN EXAMINATION OF DATA

4.1.   State Taxation of Alcoholic Beverages

In an effort to see if regulation theory sheds light on taxation as regulation, we examined state taxes on three inelastically demanded goods—beer, wine, and liquor as measured in dollars per gallon for each of the 50 states and the District of Columbia. Each state imposes taxes on beer and wine, and each state either taxes liquor or sells its through state-owned outlets (Wagner, 1997). We make a simple statistical examination of tax and sales outlet data and draw inferences that allow us to assign states to three categories: (1) strict Pigouvian, (2) special interest motivated, and (3) undefined, which may simply mean they are revenue maximizers. Our classification scheme is based on an analysis of outliers. We labeled as outliers all states that had taxes higher or lower than the all-state average by more than one standard deviation.

A strict Pigouvian theory predicts that a state that chooses to place signifi-
cantly higher (lower) taxes on one alcoholic beverage will do the same for all
beverages. Consumption of alcohol whether from beer, wine, or liquor is the
socially harmful activity to be reduced. A special interest theory predicts that
state taxation will reflect a multitude of issues including those raised by the
"anti-sin" lobby. In some cases, the pro-beer lobby will win out; in other states,
wine producers will take the day. Under this theory, there can be significantly
higher state taxes for one beverage than the average for all states while another
beverage is taxed at either a lower level or about the same as the average for
the nation. The differential effects shuffle demand across product categories,
giving competitive advantages and disadvantages along the way. The unde-
fined or revenue maximizing states will simply tax all beverages at levels that
approximate the all-state mean. We report our findings in the accompanying
Table 1. States that satisfy the Pigouvian criteria are in bold face type.

TABLE 1.
States with Alcoholic Beverage Taxes Significantly
Higher or Lower than the All-State Average

State	Beer	Wine	Liquor
Alabama	**Positive**	**Positive**	**State outlet**
Alaska			Positive
California		Negative	
Florida	**Positive**	**Positive**	**Positive**
Georgia	Positive	Positive	
Hawaii	**Positive**	**Positive**	**Positive**
Iowa		Positive	State outlet
Louisiana		Positive	
Kentucky			Negative
Maryland			Negative
Mississippi	Positive		State outlet
Missouri	Negative		Negative
Nevada			Negative
New Mexico		Positive	Positive
New York		Negative	Positive
Oklahoma			Positive
North Carolina	Positive		State outlet
South Carolina	Positive		
Texas		Negative	
Virginia		Positive	
Wisconsin	**Negative**	**Negative**	**State outlet**
Wyoming		Negative	
District of Columbia			Negative

Note: The January 1, 2000 data used for calculations made by the
authors are from http://www.taxadmin.org/ftal/rate/tax_structure.

Of the 50 states and District of Columbia, there are 23 that have at least one beverage taxed at a level that is more than one standard deviation above or below the national mean. (We note that a normal distribution would have 17 units divided equally in the two tails of the distribution.) Of these, only three states—Alabama, Florida, and Hawaii—meet the strict Pigouvian standard. These three impose significantly higher taxes on all three beverages. By our test, they do not discriminate among beverages. (We recognize the possibility that Florida and Hawaii may see alcohol taxation as a way to obtain revenue from tourists.) Wisconsin, which taxes at a significantly lower level than the nation, also behaves consistently in its treatment of alcoholic beverages. But we suspect that the behavior there reflects the efforts of a strong brewer and distiller lobby, not a belief that there are positive spillover effects that need to be encouraged.

4.2.   A Focus on Differential Effects

What can be said about the 20 outlier states in the table that discriminate among alcoholic beverages and tax at levels that differ significantly from the all-state average? We note the large number of cases where one beverage or another is either taxed favorably or penalized and call particular attention to California, which seems responsive to the wine lobby, and to Kentucky, which apparently favors the bourbon producers. Obviously, the imposition of a higher tax on beer or wine causes liquor to be a relatively more attractive substitute product. The presence of differential effects is tentative evidence of successful special interest lobbying of the sort Pigou warned about.

Recall that the tax is levied on a per gallon basis, not on the basis of alcohol content. If alcohol is the desired ingredient, then a higher per gallon tax on liquor imparts a heavier burden on alcohol consumers than a similar tax on beer or wine. Substitution effects that are induced can lead to an actual increase in the level of alcohol consumed.

This effect was reported Adam Gifford (1997) in his analysis of the relative effects of prohibition on the market shares of beer, wine, and spirits in per capita consumption. Gifford reports that the consumption of spirits and wine actually increased during prohibition, while the consumption of beer dropped dramatically. He explains that the cost of monitoring shipments of beer in barrels was much lower than the cost of monitoring liquor and wine shipments. Prohibition imposed different effective taxes on the three beverages. The accompanying table, taken from Gifford, shows consumption of various beverages and the pure alcohol equivalent in gallons per capita prior to prohibition, immediately after the passage of prohibition, and later following market adjustments. As indicated in Table 2, the total intake of pure alcohol fell by a bit more than 30 percent during prohibition, but beer production, which was dominated by local breweries at the time, was practically wiped out.

TABLE 2.
Prohibition Related Alcoholic Beverage Consumption
Average Annual Gallons per Capita

Period	Spirits	Beer	Wine	Total Pure Alcohol
1911–1914	1.47	20.53	0.59	1.69
1921–1922	0.92	1.49	0.51	0.73
1927–1930	1.62	6.27	0.98	1.14

Source: Gifford, Adam, Jr. (1997). "Whiskey, Margarine, and News-papers: A Tale of Three Taxes." *Taxing Choice*, Shughart II, William F., ed., p. 66. New Brunswick, NJ: Transaction Press.

Gifford notes that the data illustrate the Alchian-Allen shipping-the-good-apples-out theorem. Good apples gain an advantage over bad ones in distant markets when freight rates are the same for both categories. With penalties for violating prohibition law about the same whether for producing and selling beer or spirits, and with the risk of detection higher for beer in barrels than booze in bags, beer took it on the chin.

4.3.   Summary

Are state alcoholic beverage taxes about closing the gap between social and private costs? Do they reflect an effort to offset the behavior of uninformed consumers? Or are they revenue sources that also reflect special interest demands within a given a state? Our examination of data suggests that alcoholic beverage taxes reflect a combination of interests, including the double-dividend interest of politicians who seek to obtain revenues while satisfying a broad "anti-sin" interest. The inconsistencies in the data, as reflected by the differential pricing across beverages, supports Pigou's concern regarding the capability of politicians to apply taxes based on his theory. We seriously doubt that state alcoholic beverage taxes systematically reflect an effort to close the gap between marginal social cost and marginal private cost. It seems to us that the special interest theory is more informative.

4.4.   Taxing Environmental Use in the OECD Countries

Our second empirical investigation focuses on the use of taxes to regulate environmental use. We used data for 21 OECD countries that reported environmentally related taxes in 1997. The United States was omitted due to its lack of environmental taxes, as were Greece, Iceland, Italy, Luxembourg, Mexico, Poland, and Turkey (OECD Environment, 2000, GAO, 1993). We return to the U.S. omission later. The annual per capita environmental tax revenue for the 21 countries we examined is shown in Table 3.

TABLE 3.
1997 OECD Environmental Taxes
Per Capita (1997 U.S. Dollars)

Country	Taxes Per Capita
Australia	1.55
Austria	608.87
Belgium	444.49
Canada	9.31
Czech. Rep.	147.23
Denmark	1,214.16
Finland	786.87
France	5.10
Germany	575.95
Hungary	88.69
Iceland	287.45
Japan	571.29
Netherlands	1,023.41
New Zealand	431.32
Norway	782.07
Portugal	360.57
Spain	306.47
Sweden	974.98
Switzerland	409.28
United Kingdom	145.19

Source: OECD primary data and calcula-
tions by authors.

While the United States rarely imposes federal taxes for environmental pur-
poses, Barde's (1997, p. 227) OECD survey reports 149 instances of environ-
mental taxes and charges with the Nordic countries, Denmark, Finland, Nor-
way, and Sweden, making the largest use of these fiscal instruments. Among
the OECD countries, environmental taxes and charges are applied to fertilizers,
pesticides, batteries, disposable razors, disposable cameras, auto oil changes,
diesel fuel, leaded and unleaded gasoline, to the production of noise, waste wa-
ter, sewage, and to carbon and sulfur emissions (OECD, 1999). A first glance
through Pigouvian eyes at this partial list of taxed activities suggests there are a
host of things that impose unwanted social costs to OECD people that need to
be internalized.[5] Viewed differently, the long list and the heavy use of energy
taxes suggest the OECD countries have identified a reliable environmental tax
base for feeding the expanding demands of government. We hope to distin-
guish between these two competing theories of environmental taxation.

Our effort to discriminate between a strict Pigouvian and some other theory
relied on a series of regression models designed to explain annual per capita
revenues from environmental taxes across OECD countries, the data displayed

in Table 3. The basic model we report took the following general form:

$$\text{ENVTAXES} = f \text{ (GDP/POP, DENSITY, ENV QUALITY,}$$
$$\text{CLEAN ENERGY),}$$

where ENVTAXES, self-reported environmental taxes per capita, is in 1997 constant purchasing parity dollars, GDP/POP is 1997 GDP per capita in dollars, and DENSITY is thousands of people per square kilometer. We expected GDP/POP to be positively related to ENVTAXES for two reasons. A first argument says that the environment is a normal good. A second argument states that ENVTAXES fund publicly provided services that grow with income. Pigouvian theory predicts that environmental taxes rise with population density, since there is greater human exposure to pollution and therefore greater social cost. If the taxes are collected for general revenue purposes, an efficient government theory suggests taxes will be lower for concentrated populations because of economies of scale in serving human communities.

ENV QUALITY entered the model by way of two different variables. We used AIR QUALITY, which was kilograms of sulfur dioxide emissions per capita, and WATER QUALTIY, which was the percent of the population having 1996 access to safe water. Our interpretation of Pigouvian theory yields ambiguous predictions for the sign on AIR QUALITY. Relatively clean air can be the result of higher environmental taxes; alternately, dirtier air will generate higher taxes. However, the theory does not allow for a zero coefficient; it predicts the coefficient will be significant. In a similar vein, the coefficient on WATER QUALITY can be either positive or negative. More access to clean water can result from higher taxes on polluters or from a clean environment that yields low tax revenue. Once again, the strict Pigouvian theory does not allow for a zero-valued coefficient. Finally, we included CLEAN ENERGY, the percent of energy provided by nuclear and hydroelectric means. All else equal, environmental taxes will be lower, the larger the share of energy accounted for by clean energy. We report two estimates in Table 4 where t-statistics are shown in parentheses. Model I contains all the variables we have described here. We deleted WATER QUALITY, which had a coefficient of almost zero significance, implying a lack of support for the strict Pigouvian interpretation, and report the results as Model II. Our discussion now focuses on Model II.

We first note that GDP/CAP has a positive sign, which supports Pigouvian and revenue growth theories, but the coefficient is significant at only the 12 percent level. In our experiments, we learned that AIR QUALITY and GDP/CAP are highly collinear. When included without AIR QUALITY, GDP/CAP is highly significant. DENSITY has a negative coefficient, but is also significant at only the 12 percent level. The negative or zero coefficient implies a lack of support for Pigouvian theory. The coefficient on AIR QUALITY, which is

TABLE 4.
OECD Environmental Taxes Regression Analysis

Variable	Model I	Model II
INTERCEPT	699.55	765.46
	(1.933)	(2.602)
GDP/CAP	0.0127	0.0127
	(1.269)	(1.640)
DENSITY	−789.69	−801.09
	(−1.508)	(−1.643)
WATER QUALITY	0.7465	
	(0.235)	
AIR QUALITY	−8.664	−9.3290
	(−2.353)	(−3.279)
CLEAN ENERGY	−10.772	−10.6348
	(−2.104)	(−2.229)
Adjusted R-Squared	0.458	0.534
$F_{(5, 14)}$	4.21	
$F_{(4, 16)}$		6.73

sulfur dioxide emission per capita, is negative and highly significant. Higher levels of sulfur dioxide emissions are associated with lower per capita environmental taxes, which could support either the Pigouvian or revenue maximizing theory. Finally, the coefficient on CLEAN ENERGY is negative and significant. When clean energy accounts for a higher share of total energy produced, environmental taxes are lower. In a sense, there is just less to tax, no matter what the motivation.

The results provide some support for a strict Pigouvian interpretation of OECD environmental taxes. The coefficient on AIR QUALITY is significant and negative. However, the results for DENSITY reject the Pigouvian theory, as do the results for WATER QUALITY in the first estimate. In further explorations, we replaced the sulfur dioxide with carbon dioxide. The results were even weaker.

4.4.1.   Summary   The OECD countries offer a rich laboratory for investigating the use of environmental taxes. Environmental taxes are widely used, and experience with them spans several decades. Can we say that the OECD experience supports the strict Pigouvian theory? Is there overwhelming evidence that environmental taxes seem applied in ways that internalize external costs? Or is the evidence mixed? Our investigation focused on relationships between all environmental taxes per capita and other economic and environmental variables. The results are mixed. We find some support for the

Pigouvian theory, but the support is not consistent for variables in our model. We note, however, that it is certainly possible that one country or a group of countries could be following Pigou's strict instructions.

In his impressive survey of the status of OECD environmental taxes, Barde (1997) indicates that it is not clear that OECD countries actually use environmental taxes to achieve environmental improvement. More to the point of Pigou's concern regarding the political economy of his namesake taxes, Barde (1997, pp. 234–235) indicates: "The 'greening' of taxes may be a political argument to justify or hide the real purpose of revenue raising." Pigou could not have stated the problem more clearly. We noted at the outset of this section that data for the United States was not available in the OECD data set. This is partly due to the fact that the United States does not impose environmental taxes on auto fuels. As indicated by Gushee and Lazzari (1993, CRS5) in their survey of federal and state motor fuel taxation, the "present structure of Federal excise taxes on motor fuels evolved from three public policy concerns: (1) revenue generation for budget deficit reduction; (2) revenue generation for highway infrastructure financing; and (3) energy policy considerations." The United States makes no pretense about fuel taxation as regulation. Our analysis of OECD data suggests that those countries might well take the same position, at least insofar as environmental regulation is concerned.

5.   FINAL THOUGHTS

We have little doubt but that taxation can be used and is used for regulatory purposes. We also have little doubt that the use of regulatory taxation is confounded by the politics of taxation. Put another way, we are convinced that Pigou was correct when he explained that politicians on the way to designing a regulatory tax would be diverted by special interests. What may begin as an efficiency enhancing tax ends as a hybrid that also enhances revenue. Of course, there is no way to determine political motivation. The sin taxes that we observe about us may simply be the result of a political effort dressed in anti-sin clothing but motivated by a search for revenue.

Putting motivation to one side, our review and analysis lead us to conclude that there are no pure Pigouvian taxes in operation today. There are simply taxes that at times are justified by stories about the need to correct for spillover costs, imperfect knowledge, and other features of life that cause marginal social cost of the taxed activities to exceed the marginal social benefits. Interestingly enough, regulatory taxes will be imposed most frequently on those activities and articles of consumption that are least affected by taxes. Taxation as regulation works so long as the taxation does not eliminate the thing being taxed.

NOTES

1. The 1971 paper was followed in 1974 by another paper that sought to identify and explain the then emerging theories of regulation (Posner, 1974). Posner's work was joined that of Becker (1983), Peltzman (1976), Stigler (1971), and combined with Buchanan, Tollison, and Tullock (1980) and Tullock (1967) to form a body of theory that contributed mightily to an enhanced understanding of regulation.

2. In his classic public finance textbook, Richard Musgrave (1959, p. 178) discusses regulatory taxation, "taxes that by definition are designed to interfere with the equal treatment of equals." Musgrave describes two categories of regulatory taxation: (1) sumptuary taxes. These are sumptuary taxes, "such as discriminatory taxes against liquor and tobacco." These are the negative counterpart of public expenditures to support "free education or school luncheons," Then, there are taxes that can improve efficiency, such as taxes that might alter the behavior of monopolists or affect the structure of industry.

3. We are obviously not talking about the "double dividend" theory that emerged in recent years with regard to environmental taxes. The argument there claimed that placing taxes on pollution will lead to a cleaner world—the first dividend—and the opportunity to reduce taxes on labor, yielding a second dividend by the elimination of the efficiency loss. (See Bohm, 1997, O'Riordan, 1997.) We note that there is always more to the story. Firms may seek regulation as an effective barrier to entry and successful cartelization. Taxes will not do this unless they discriminate against new entrants. (See Buchanan and Tullock, 1969.)

4. Gravelle and Zimmerman (1994) report estimates of price elasticity of demand for cigarettes. The short-run average for all consumers has an absolute value of 0.31 as compared to the long-run of 1.20. These vary markedly for different age groups, with the short run elasticity of older smokers being 0.15, as compared to teenagers' elasticity of 1.2. Schumacher and Fried (2000) report the results of the Alaskan experience with a large increase in tobacco product taxes are consistent with an elasticity estimate of 0.40. The absolute values of price elasticities of demand for alcoholic beverages reported by Grossman, Sindelar, Mullahy, et al. (1993) are 0.30 for beer, 0.70 for wine, and 0.80 for spirits. The OECD (2000) estimates for gasoline, which is highly taxed for apparent conservation and environmental reasons, range from 0.15 to 0.67 in the short run to 1.25 in the long run.

5. We note the extensive debate that still rages when a Pigouvian approach is suggested for environmental control purposes. For a discussion of the pros, cons, and alternatives, see Backhaus (1998), Barnett (1980), Baumol (1972), and Yandle (1999).

REFERENCES

Backhaus, Jurgen. (1998). "When Should the Ecotax Kick In? A Cost and Choice Approach." *Kyklos* 51: 565-575.

Barnett, A. H. (1980). "The Pigouvian Tax Rule Under Monopoly." *American Economic Review* 70 (December): 337-410.

Baumol, William J. (1972). "On Taxation and the Control of Externalities." *American Economic Review* 62 (June): 307-322.

Becker, Gary S. (1983). "A Theory of Competition Among Pressure Groups for Political Influence." *The Quarterly Journal of Economics* 118 (August): 371-399.

Barde (1997).

Bohm, Peter, (1997). "Environmental Taxation and the Double Dividend: Fact or Fallacy?" *Ecotaxation,* O'Riordan, Timothy, ed., pp. 106-124. New York: St. Martin's Press.

Buchanan, James M., Tollison, Robert D., and Tullock, Gordon. (1980). *Toward a Theory of the Rent-Seeking Society.* College Station, TX: Texas A&M Press.

Buchanan, James M. and Tullock, Gordon. (1969). "Polluters' 'Profit' and Political Response." *American Economic Review* 59 (March): 174-177.

GAO (1993). Environmental Protection: Implications of Using Pollution Taxes to Supplement Regulation, GAO/RCED-93-13. Washington, DC: U.S. General Accounting Office. (February).

Gifford, Adam, Jr. (1997). "Whiskey, Margarine, and Newspapers: A Tale of Three Taxes." *Taxing Choice: The Predatory Politics of Fiscal Discrimination,* Shughart II, William F., ed., pp. 57-77. New Brunswick, NJ: Transaction Press.

Gravelle, Jane and Zimmerman, Dennis. (1994). "Cigarette Taxes to Fund Health Care Reform." *National Tax Journal* 47 (Sept.): 575-591.

Grossman, Michael, Sindelar, Jody L., Mullahy, John et al. (1993). "Policy Watch: Alcohol and Cigarette Taxes." *The Journal of Economic Perspectives* 7 (Autumn): 211-222.

Gushee, David E. and Lazzari, Salvatore. (1993). *Disparate Impacts of Federal and State Highway Taxes on Alternative Motor Fuels.* Washington, D.C.: Congressional Research Service. Library of Congress (March 12).

McChesney, Fred S. (1987). "Rent Extraction and Rent Creation in the Economic Theory of Regulation." *Journal of Legal Studies* 16 (January): 101-118.

Musgrave, Richard A. (1959). *The Theory of Public Finance.* New York: McGraw-Hill, Inc.

OECD (1999). Economic Instruments for Pollution Control and Natural Resource Management in OECD Countries: A Survey. ENV/EPOC/GEEI(98)35/REV1/FINAL. Paris (Sept. 24).

OECD (2000). Behavioral Response to Environmentally-Related Taxes, COM/ENV/EPOC/DAFFE/CFA(99)111/FINAL. Paris (March).

OECD Environment. (2000). Environmentally Related Taxes Database, http://www.oecd.org/env/policies/taxes/index.htm (July 7).

O'Riordan, Timothy. (ed.) (1997). *Ecotaxation.* New York: St. Martin's Press.

Peltzman, Sam. (1976). "Toward a More General Theory of Regulation." *Journal of Law & Economics* 19 (August): 211-240.

Pigou, A. C. (1938). *Economics of Welfare,* 4th edition. London: Macmillan & Company, Ltd.

Pigou, A. C. (1960). *A Study in Public Finance.* London: Macmillan & Company, Ltd.

Posner, Richard A. (1971). "Taxation by Regulation." *Bell Journal* 2 (Spring): 22-50.

Posner, Richard A. (1974). "Theories of Economic Regulation." *Bell Journal* 5 (Autumn): 335-358.

Ramsey, Frank P. (1927). "A Contribution to the Theory of Taxation." *Economic Journal* 37 (March): 47-61.

Schumacher, Catherine and Fried, Brett. (2000). The Impact of the 1997 Tobacco Tax Rate Increase in Alaska. http://www.hss.state.ak.us/tobacrep.htm, 7/13/00.

Shughart, William F. II. (ed.) (1997). *Taxing Choice: The Predatory Politics of Fiscal Discrimination.* New Brunswick, NJ: Transaction Press.

Sirico, Robert A. (1995). *The Sin Tax.* Grand Rapids, MI: The Acton Institute.

Stigler (1971).

Tullock (1967).

Wagner, Richard. (1997). "The Taxation of Alcohol and the Control of Social Costs." *Taxing Choice: The Predatory Politics of Fiscal Discrimination,* Shughart II, William F., ed., pp. 227-246. New Brunswick, NJ: Transaction Press.

Yandle, Bruce. (1983). "Bootleggers and Baptists: The Education of a Regulatory Economist." *Regulation* 7: 12-15.

Yandle, Bruce. (1999). "Bootleggers and Baptists in Retrospect." *Regulation* 22: 5-7.

Yelvington, Brenda (1997).

Chapter 10

TAXATION, BLACK MARKETS, AND OTHER UNINTENDED CONSEQUENCES

Carla Marchese*
University of Eastern Piedmont
carla.marchese@unipmn.it

Abstract   The economics of tax evasion and tax avoidance is shortly presented, and the relevant literature is surveyed. Starting from the income tax and the standard portfolio choice model, the perspective is enlarged by adding further circumstances (e.g., labor supply, indirect taxation) and by studying optimal enforcement policies. To understand the taxpayer's behavior, the contribution of new theoretical perspectives (e.g., first order risk aversion, network externalities) is examined. Topics pertaining to moral, social and political implications of tax evasion are considered.

Keywords: tax evasion, tax avoidance, enforcement

JEL classification: H26, D82, K34

1.   INTRODUCTION

This chapter examines unintended consequences of taxation, other than excess burden. These are primarily tax avoidance and tax evasion.[1] In specific circumstances, the coercive power of the state exerted through taxation can be frustrated by taxpayers who exploit loopholes in the law, cheat on tax declarations, conceal income etc. Tax evasion is sometimes associated to other illegal activity, such as bribery and extortion, and is ordinarily practiced in illegal markets. The latter circumstances imply free riding with reference to state financing; however when illegal conduct takes the form of organized crime, some peculiar form of alternative collective good provision may also arise.

Quite different is the case wherein the refusal to pay taxes forms part of a radical political protest, and is the prelude to revolution and the attempt to form an alternative political order.

A vast body of literature has studied evasion of income tax as an individual choice under risk. The paradigm adopted in this case is that of portfolio allocation: a utility maximizing taxpayer can invest part of her given budget into the risky activity of tax evasion. Alternatively, under the same assumptions, paying taxes can be viewed as a form of insurance against the risk of audits and sanctions. Section 2 presents this basic model, which furnishes predictions concerning the effects on tax evasion of parameters such as probability of control, sanctions and tax rates. In Sections 3 and 4 the basic model is enlarged to endogenize choices concerning labor supply and occupational choice. When the taxpayer's income is considered as endogenous, tax evasion may in some cases prove to be of social benefit rather than detriment, as it could foster risk taking and income growth.

Tax evasion is fought by states at a cost through enforcement policy. Assuming that the state can commit to a given enforcement strategy, a principal agent model can be used to formulate a revenue maximizing taxation and audit policy (Section 5). On the other hand, if commitment is not possible, a Nash equilibrium between two players (the enforcing agency and the taxpayer) can be devised (Section 6). Sanctions are an important incentive that can be adjusted to foster compliance: Section 7 draws from law and economics to discuss their role.

The basic model presented in Section 2 has been criticized on many grounds. The choice whether to pay taxes or to evade them gives rise to a moral problem, and implies social interactions in which psychological factors may play an important role (Section 8). Moreover, the description of individual choice under risk provided by the expected utility approach can be amended to better cope with empirical evidence (Section 9).

Optimal taxation models (Section 10) have treated tax evasion in a manner which parallels that of excess burden; that is, as a limiting factor with the potential for reducing the advantages of attempts at collecting revenue in order to supply public goods and to redistribute income. Whether this is a good or an evil is determined by the ultimate goals of the state. If Buchanan's Leviathan is substituted for the benevolent dictator, the check stemming from tax evasion may exert beneficial effects (Section 11).

Applied public choice emphasizes rent-seeking as an explanation for both the demand for tax evasion (by citizens and interest groups), and the supply of lenient policies toward it (by politicians and bureaucrats). To prevent rent-seeking, which aims at preferential treatment in terms of avoidance and evasion opportunities, constitutional clauses concerning generality of taxation, auditing and sanction policies can be introduced (Section 11).

Tax evasion may involve network externalities between evaders: a model of smuggling that uses this approach is presented in Section 12. Section 13 examines tax avoidance, which, in contrast to tax evasion, refers to non risky activity intended to reduce the tax burden at a cost. Conclusive observations are reported in Section 14.

2.   THE BASIC MODEL

Much of the literature on tax evasion is based on the Allingham and Sandmo (1972) model, which focuses on proportional income taxation and describes taxpayer choice concerning the amount of income to report as a decision under uncertainty. Let us consider a risk averse taxpayer, who faces two possible states of the world. If she is not audited, her disposable income is:

$$Y = W - tX \tag{1}$$

where W is the taxpayer's true income, t is the tax rate and X her reported income. If she is audited, her disposable income is:

$$Z = W - tX - F(W - X) \tag{2}$$

where $F > t$ is a penalty rate. It is assumed that sanctions apply to unreported income, and that audits always reveal true income. By solving (1) for X and substituting the result into (2) the taxpayer's budget constraint is obtained:

$$Z = \left(\frac{F}{t} - F\right)W - \left(\frac{F - t}{t}\right)Y.$$

The slope of the taxpayer's budget line (see Figure 1) is $- [(F - t)/t]$. While a faithful taxpayer has a disposable income of $(1 - t)W$ in both states of the
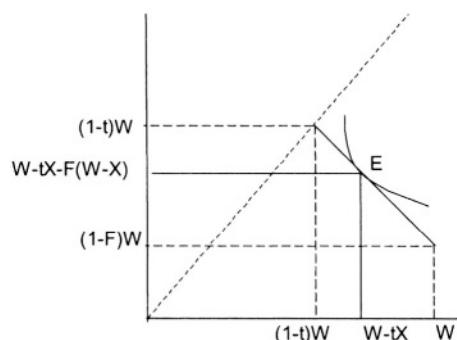
FIGURE 1. The Taxpayer's Equilibrium

world (she remains on the 45° degree line, the so called "certainty line"), a full evader has a disposable income of W if she is not caught, and of $(1 - F) W$ if she is caught.

Assuming that the taxpayer is amoral and unaffected by any perspective of social stigma, her problem is simply the maximization of her expected utility:

$$(1 - p)U(Y) + pU(Z)$$

where p is the probability of detection. The taxpayer's equilibrium (with an interior solution: see Figure 1) is reached when:

$$\frac{(1 - p)U'(Y)}{pU'(Z)} = \left(\frac{F - t}{t}\right) \tag{3}$$

where the prime is used to indicate the first derivative. The left hand side of equation (3) refers to the taxpayer's marginal rate of substitution between income in the two states of the world, while the right hand side indicates the slope of the budget line.

In Figure 1 the paid tax is represented by the distance $W - (W - tX)$, while the evaded tax is represented by the distance $(W - tX) - (1 - t) W$.

As a full complier has an equal marginal utility of income in both states of the world, with full compliance (3) boils down to:

$$\frac{(1 - p)}{p} = \left(\frac{F - t}{t}\right)$$

which is satisfied if $pF = t$. Thus full compliance occurs whenever the tax is equal to or less than the expected sanction.[2]

Within this model a comparative static analysis can be performed with reference to the key parameters F, p, α, t, W, where α measures the taxpayer's attitudes toward risk (usually in terms of either absolute or relative risk aversion). Raising the sanction F gives rise to two effects (Figure 2): on the one hand, the slope of the budget line increases in absolute value. On the other, the marginal rate of substitution, for every reported income X, diminishes, as the marginal utility of income of a detected evader $U'(Z)$ is now higher than before. These facts imply that the taxpayer's equilibrium must occur at a point such as E' in Figure 2, with lower tax evasion.

A similar consequence stems from an increase in the detection probability. In this case the budget set does not change, whereas the indifference curve slope diminishes in absolute value (as $(1 - p)/p$ is lower: see equation (3)). Thus, equilibrium occurs at a point to the left of the previous one (characterized by lower tax evasion).
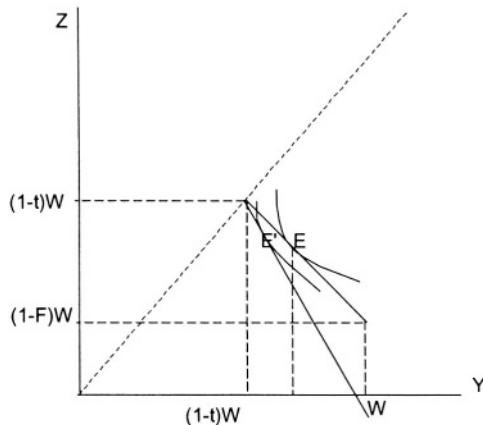
FIGURE 2. The Effect of Increasing the Penalty Rate.

Increasing the penalty or the probability of detection thus has clear cut effects, in line with intuition: raising deterrence has beneficial consequences against tax evasion. Instead, increasing the tax rate does not have such clear cut effects. At a higher tax rate, the disposable income of a faithful taxpayer is lower than before, while there is no impact on a full evader. The absolute value of the slope of the budget constraint is thus lower than before. There is a substitution effect that favors tax evasion, and an income effect, whose sign depends on the attitudes toward risk. To establish the sign of the income effect let us assume, as it is often held, that absolute risk aversion decreases as income increases (from now on: DARA, Decreasing Absolute Risk Aversion). A tax rate increase, negatively affecting taxpayer's income in both states of the world at every level of reported income, should thus induce a decrease in tax evasion. Consequently, the substitution and the income effects run one against the other, and the final effect depends on which one prevails. The problem of the effects of tax rates changes on compliance, however, is controversial. Other models, in fact, produce different results: we will return to this topic at the end of this Section.

As far as risk aversion is concerned, a higher absolute risk aversion implies *ceteris paribus* a lower tax evasion, as a more risk averse person[3] has a flatter indifference curve, being less keen to renounce income in the bad state of the world (when an audit occurs) for income in the good state (no audit).

Changes in income on the one hand cause a parallel shift of the budget line, while, on the other, they influence the curvature of the indifference curve through the effects of the income level on attitudes toward risk. A very straightforward case to analyze graphically is that of CRRA (Constant Relative Risk Aversion). In this case taxpayer equilibrium points all lie on the same ray from the origin, which means that the proportion of evaded tax over the owed one is

constant. The proportion will instead increase or decrease with increasing or decreasing relative risk aversion.

While a large body of literature is based on the Allingham and Sandmo (1972) model, it is however often amended according to the suggestion of Yitzhaki (1987), who assumed that the penalty refers to the evaded tax rather than to the unreported income. The taxpayer's income in case of detection becomes:

$$Z = W - tX - (1 + s)t(W - X)$$

where $s > 1$ is the penalty rate. Following the approach of Yitzhaki (1987), which in fact is in line with legal provisions in many countries, has important consequences on the effect of tax rate increases. As the penalty also increases with the tax rate, the full evader is negatively impacted (as he would face a higher penalty in case of detection), the budget line slope does not change and thus there is no substitution effect in favor of tax evasion. Only the income effect matters. As we are considering a taxpayer whose income has been reduced by a higher tax, under DARA the taxpayer should decrease tax evasion.

The approach of Yitzhaki (1987) suggests that tax cuts may have perverse effects upon tax compliance. Common sense seems to indicate the opposite. The question has been tackled from one angle by building richer models to describe taxpayer's behavior, and from another by collecting and analyzing more ample empirical evidence in order to shed light on the question (Clotfelter, 1983, Feinstein, 1991).

3.   TAX EVASION AND LABOR SUPPLY

The taxpayer's problem becomes highly complex if income is not assumed as exogenous, but rather can vary according to the occupational choice and the number of hours worked. To render the problem tractable, simplifying assumptions are in order.

A drastic simplification consists in assuming that there is a two stage decision process. The first one refers only to the total number of hours to work, without contemplating the possibility of evading taxes. At this stage the fiscal system influences the choice only through the tax rate, as in standard labor supply models. At the second stage the choice is about the amount of income to be reported: in this case the choice is made under uncertainty, and the probability of detection and the amount of the sanction are influential. One can also interpret the second choice as a decision about how many hours to work in the legal market (whose earnings will be reported) and how many in the black economy, out of a fixed number of labor hours chosen in the first stage, and provided that the gross wage rate is the same in both sectors.[4]

The comparative static analysis of the effects of changes in fiscal parameters gives the same results as before for the parameters p and F, as they matter only at the second choice stage. To examine what happens at the first decision stage, let us assume that the tax system is progressive and characterized by a linear negative income tax: each taxpayer has a constant marginal tax rate t and receives a lump-sum transfer A. The effects of increased progressivity of the tax system upon labor supply are depicted in Figure 3.

The taxpayer budget line increases in the number of hours worked L (that is $Y = A + (1 - t)wL$). Indifference curves have a positive slope, as labor is considered a "bad," and refer to two agents representative of the poor and the rich respectively. The flatter budget constraint refers to a more progressive tax schedule. If a more progressive system is introduced, for the poor, who qualify for a net transfer, the labor supply diminishes (the equilibrium point shifts from E_p to E'_p), whereas disposable income Y increases. Let us now examine the consequences upon tax compliance (Figure 4). On the one hand the budget line shifts upward (as the taxpayer is richer), on the other—due to the larger marginal tax rate—it flattens (there is a substitution effect in favor of tax evasion). Thus, assuming DARA, tax evasion (or moonlighting) increases. The same effect also arises when sanctions are applied on the evaded tax. In
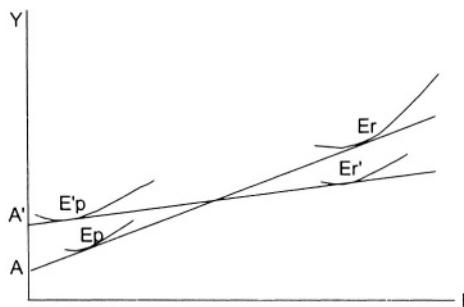
FIGURE 3. Progressive Taxation and Labor Supply.
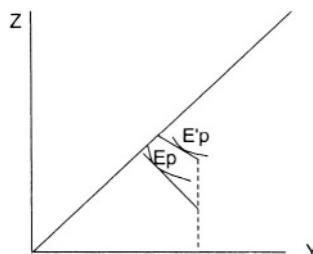
FIGURE 4. Progressivity and Tax Evasion.

this case the slope of the budget line does not change, but the income effect is enough to produce the result under DARA.

Let us now consider the rich, who, according to Figure 3 have to pay a net tax (the equilibrium point shifts from Er to Er'). Whether she chooses to reduce the number of hours of work (as in Figure 3), or whether she increases them, her net income should fall.[5] The consequences upon tax evasion are thus ambiguous, as there is a substitution effect in favor of tax evasion, and an income effect caused by the impoverishment of the taxpayer, which, under DARA, should raise compliance. This result, in fact, parallels that which is found with exogenous income, although here the income effect is reduced in the case of backward bending labor supply, and reinforced in the opposite case (considered in Figure 3). If sanctions are charged upon the evaded tax, the ambiguity is solved under DARA in favor of higher compliance of the rich, as only the income effect matters. Thus, summing up, higher progressivity implies greater tax evasion at low income levels, and ambiguous results at higher ones.

Participation in the black labor market could also be encouraged when the number of hours of work is constrained by an upper-bound in the legal market. If hours in the black economy are perfect substitutes, workers will try to circumvent the limit by participating in the black market. When the upper constraint on hours of work in the legal market is binding, a tax rate increase always increases the number of hours worked in the black market, as it affects legal work which is inframarginal, and thus produces no substitution effect, and only income effect. Income reduction, when leisure is a normal good, implies a larger work supply, which shows up in the illegal market (Lacroix and Fortin, 1992).

Empirical analysis based on questionnaires (Lacroix and Fortin, 1992) indicates that tax rate increases induce those who participate in both sectors to increase hours worked in the irregular one, while penalties and probability of detection have the opposite effect. Irregular hours are also highly elastic with reference to the wage in the regular sector.

4.   TAX EVASION AND OCCUPATIONAL CHOICE

It is widely held that evasion opportunities are not evenly distributed in the economic system. For example some activities are more visible than others and so the income they produce is less evadable. Underlining this fact, many models assume a two sector economy and structure the agent's choice in two stages, by assuming that at the first stage the agent chooses the sector, while at the second stage either he complies with certainty (if he has chosen the non evadable sector) or he decides whether to comply or not (if he has chosen the evadable sector). Key features within this approach are:

1. The agent's remuneration within each sector (which could either be influenced or not by the number of people who make the same choice, according to assumptions about competitiveness of the markets and technological parameters).
2. Idiosyncratic factors (such as abilities, risk aversion etc.) that drive the choice toward one specific sector.
3. Labor productivity in each sector. Some authors assume that the evadable sector is more productive (e.g., since participants are more prone to risk bearing), others, instead, focus on the fact that it wastes resources in concealing income and assume that it is less[6] productive[7].

Pestieau and Possen (1991) assume a fixed labor supply, while occupation can either be risky (e.g., the taxpayer becomes an entrepreneur) or sure (the taxpayer works as an employee). The risky perspective implies the possibility of cheating, as the state cannot discover (except by auditing the taxpayer) whether he earned a high income b or a low income a. On the contrary, the taxpayers who choose the sure occupation are fully visible, and their income c (lower than the expected value for entrepreneurs) is taxed with certainty without any audit. Let us assume that the remunerations $a, b, c,$ as well as the probabilities of occurrence of either a or b are fixed and not influenced by the number of people who choose either type of occupation. Assuming that there is no barrier to entry in either sector, equilibrium is reached if each agent's expected utility in the risky sector equals his utility from the sure occupation. To structure the model, Pestieau and Possen (1991) assume that all taxpayers are characterized by CRRA, while they differ with reference to the relative risk aversion coefficient. Within a model that assumes a linear negative income tax, random audits, and a fixed penalty rate charged upon undeclared income, people characterized by risk aversion higher than a given threshold choose the sure occupation, while the less risk averse ones become entrepreneurs. Moreover, if the penalty is not too high, entrepreneurs separate into evaders (the least risk averse) and compilers.

With reference to tax parameters, a clear cut result is reached for the lump sum transfer provided by the negative income tax. In a similar vein to that of the model examined in Section 3, raising the transfer unambiguously increases tax evasion[8], but also pushes people toward the more productive role of entrepreneurs.

Within the Pestieau and Possen (1991) approach the deadweight loss associated with policies aimed at reducing tax evasion becomes visible, as increases in the rate of audits on the one hand reduce tax evasion, while on the other they drive people toward sure occupations (with a lower expected income). There is thus a peculiar kind of "Laffer curve" problem in collecting revenue, as compliance grows while the tax base shrinks.[9]

5.   ENFORCEMENT

Let us now consider how the Tax Administration can design the enforcement of the tax law. This problem is part of the more general problem of designing the tax system itself. To keep things simple, for the moment we will set aside the equity problems that this design implies, and assume that the Tax Administration simply aims at maximizing expected net revenue. The problem is not trivial, as the Tax Administration can verify each taxpayer's income only by auditing him at a cost, while the taxpayer knows his true income. Usually taxpayers are asked to send an income report, which, however, may be not accurate. The Tax Administration problem is somehow eased if, as we shall assume, it can commit to an audit strategy, that is, it credibly announces in which circumstances a taxpayer will be audited. Each audit costs c. Let us also consider, following Reinganum and Wilde (1985), risk neutral taxpayers[10] who cheat whenever the expected benefits of cheating are positive (see also Section 2). Each taxpayer's income is a random variable, independently and identically distributed. We shall consider specifically a uniform income distribution over the support $[0, Y\max]$.

The problem of setting taxes, fines and the audit rate can be formulated through a principal agent model, where the principal (the Tax Administration) provides incentives in order to induce the agent to reveal his income.

Let us consider an extremely simplified tax system. There is only a lump sum tax T, that must be paid by the taxpayers who report an income higher or equal to T. Taxpayers who report a lower income, must surrender their reported income. Moreover, if an audit reveals that they cheated, they must supplement their payment and pay a fine whose absolute amount is θ, so that their total payment eventually amounts either to $T + \theta$, or to their true income if the latter is lower than $T + \theta$.

Let us assume that the Tax Administration audits randomly and sets the audit probability without considering taxpayer's reports. In this case the Tax Administration problem has two simple solutions. It should set the probability to 1, auditing all the taxpayers, while setting the tax $T = Y\max$, if the audit cost c is lower than the average income; or it should set the probability to 0, renouncing auditing altogether and raising no revenue, if c is higher than the average income. By auditing everyone, with the tax set to $Y_{\max}$, it is possible to collect the whole national income, as everyone has to surrender his true income. Hence, if the audit cost is lower than average income, net revenue is maximized. If however the audit cost is higher than average income, auditing would cause a loss.

Better results can be achieved by a cut-off rule. The Tax Administration commits to auditing each taxpayer who reports an income lower than a threshold i^*. In this case the optimal policy is to set $i^* = T$,[11] while choosing T

in order to maximize net revenue. Net income is maximized by setting i^* $= T = Y_{max} - c$ (In Figure 5 Y max E = c).

To illustrate the functioning of this rule, let us assume that there is a continuum of taxpayers, ordered from the poorest to the richest, with an index $i = 0 \ldots N$. In Figure 5 the taxpayer's index is reported on the abscissa and the corresponding income on the ordinate. National income is measured by the triangle $0Y_{max}N$; there are as many taxpayers as income levels, so that the line $0Y_{max}$ is 45° degree. To find the optimal cut-off level, we raise it above zero until the marginal cost equals the marginal benefit. As marginally raising the cut-off level results in the need for further auditing, the marginal cost of this policy amounts to the marginal audit cost c; the marginal benefit amounts to one extra dollar of revenue, times the number of taxpayers whose income is higher than the threshold, who must raise their report for benefiting from the higher cut-off. In Figure 5 the number of these taxpayers at i^* is given by DE, which is equal to $Y_{max}E$ by construction. Figure 5 also shows that the net revenue is higher with the cut-off rule than with random auditing; audit costs decrease by the area $ABNC$, while the revenue lost amounts only to the triangle $DY_{max}E$.

In order for the cut-off rule to maximize Tax Administration revenue, the income distribution must possess specific (even if not too demanding) characteristics: that is, it must be such that the revenue loss due to the allowed evasion at the top income levels is lower than the benefit due to the reduction in audit costs.[12] As far as the influence of other parameters is concerned, the cut-off level is higher the higher the maximum income level, and the lower the audit cost c.

A justification for the cut-off rule can be found by considering that minor offenses should generally not be investigated, in order to save costs.[13] Refer-
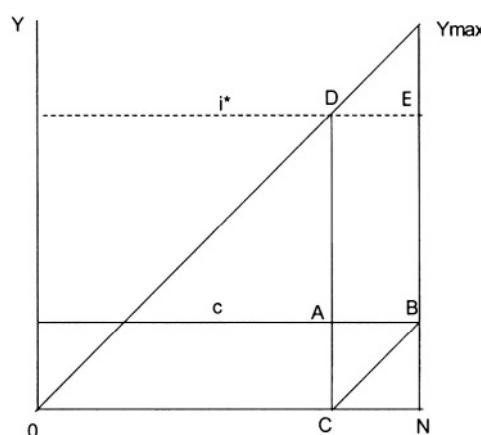
FIGURE 5. The Cut-off Rule.

ring to Figure 5, if an income report meets the cut-off income level, the offence cannot be higher than $Y_{max} - i^*$, and renouncing investigation is beneficial in terms of avoiding audit costs.

Turning to cases in which taxes are not so high as to fully expropriate tax-payers income, let us assume that the law provides for a proportional income taxation at rate $t < 1$, and for a penalty rate (on evaded income) F. In this case the Tax Administration will never set the audit probability at a level higher than t/F, which ensures full compliance. The audit probability thus is < 1, as long as $F > t$. Moreover, the Tax Administration could still pursue revenue max-imization by resorting to the cut-off rule. The cut-off level should be set at a value i^{**}, which can be determined again by equating the marginal cost of rais-ing the cut-off level (c times the audit probability t/F), to the marginal benefit in terms of extra tax payment, i.e., $t(Y_{max} - i^{**})$, where the term in parenthe-ses refers to the number of taxpayers whose income is higher than i^{**}. Hence $i^{**} = Y_{max} - (c/F)$. An increase in the penalty rate reduces the probability needed to ensure compliance, and raises the cut-off income level.

The cut-off approach relies on the assumption that the Tax Administration can audit each taxpayer. If however some taxpayers do not file any report at all, they become "ghosts," who are difficult to uncover. In fact, conditioning audits upon reports may encourage this type of response. In addition to report investigation according to the cut-off rule, some monitoring activity which is not conditional on reports may then become necessary (Mookherjee and Png, 1992; Cowell and Gordon, 1995). Signals pertaining to the living standard could be exploited (Yaniv, 2003).

The cut-off rule ensures horizontal equity, as taxpayers who submit the same report face the same audit probability. Keeping in mind that people who have the same income always report either their true income or the cut-off level, horizontal equity is also ensured with reference to true incomes. How-ever the approach considered introduces a regressive bias in auditing policy, as audits stop when the cut-off income level is crossed.

A regressive bias is often considered an unavoidable characteristic of en-forcement policies. Let us refer again to a case in which there is a propor-tional income tax t, and a penalty rate F is applied to unreported income if detection occurs, whereas the audit probability p depends now on reported in-come r. A risk neutral taxpayer for whom $p(r = W)F < t$, where W is his true income, might be willing to understate his income. If the audit proba-bility increases with reported income, he would fare better the lower is his reported income, as both taxes paid and expected penalties would decline. If however the audit probability decreases with reported income, he would face a higher audit probability the lower his reported income. Eventually, at some report level the audit probability would become high enough that he would not be willing to further underestimate his income (as in fact happens with the

cut-off rule): thus the Tax Administration would fare better with the latter approach. In fact, if we consider the taxpayer's true income as information which is hidden from the tax authority, the typical adverse selection problem induces every taxpayer to pretend to be poor. Monitoring effort must thus focus on low income reports.[14]

The regressive bias can be overcome and vertical equity can sometimes be reached if one is prepared to sacrifice horizontal equity. If the Tax Administration relies on indicators other than the income report, which convey further information about the taxpayer's income, it can form audit classes, each one characterized by a specific cut-off income level. This approach, under given conditions (Scotchmer, 1987), ensures that each income level expects to underpay a fixed amount of taxes, and thus payments are an increasing proportion of income. Equity issues in this case, however, may arise with reference to the income indicator considered, since typical indicators, such as skills, economic sector, geographical zone, age, etc., may be more or less reliable at different income levels. The latter problem also characterizes presumptive taxation, which avoids the problems stemming from tax evasion by assigning incomes and taxes by relying only upon indicators other than income.

The agency approach to tax evasion implies an incentive problem for the Tax Administration. Under a cut-off rule, audited taxpayers are always found compliant, while evaders are never audited. The Tax Administration would thus be tempted to revise the announced policy after receiving taxpayers' reports which reveal (by induction) taxpayers income. Moreover, when taxes do not expropriate the whole income, innocent taxpayers who are audited might fear that audit results are forged in order to extract additional resources from them (Border and Sobel, 1987). Taxpayers who reported the cut-off income may fear that they will be audited notwithstanding the announced renunciation, as it is apparent that they are evaders. Only if we assume that the Tax Administration can credibly commit to an audit rule will the announced policy actually be carried on, despite the fact that it does not produce any penalty revenue and is not optimal from an *ex-post* point of view.

If the relationship between the taxpayer and the Tax Administration lasts over time, audit probabilities and penalties could be made conditional on the taxpayer's past record (Greenberg, 1984; Harrington, 1988). If the taxpayer has already been found guilty of tax evasion, she should be put under special surveillance, in a group where audit probability and penalties are higher. However, she could return to the ordinary regime if a new audit reveals compliance. Let us assume infinitely living agents; the Tax Administration forms taxpayers' classes based on the agents' past performance. Audits within a class of "bad people" have a leverage effect, as taxpayers under special surveillance must trade off the benefits of evasion against the expected penalty (as usual) but also against the opportunity cost of not moving into a less guarded class.

A typical feature of this approach (which resembles results found in agency models) is that in equilibrium people under special surveillance always comply, while people in the "good class" evade. The mechanism achieves its best results if there is a closure class in which audit occurs with probability 1, and in which the penalty is fixed high enough to render tax evasion not worthwhile. This class should be empty in equilibrium, but its existence introduces a strong threat that highly favors compliance. The performance of this type of mechanism is influenced by the taxpayers' discount rate. Best results are achieved if taxpayers do not discount at all, so that the threat of spending the future in an unpleasant position has its maximal effect. The main drawback of this approach lies in the possibility of auditing mistakes. If innocent taxpayers are occasionally mistakenly found guilty of evasion, the accumulation of such cases in the long term could eventually put all the taxpayers in the class with no way out.

6.   THE TAX COMPLIANCE GAME

To avoid the odd consequences of the cut-off rule, enforcement policy can be modelled without assuming that the Tax Administration is committed to a given audit policy. Instead, it is assumed that the Tax Administration decides *ex-post* and audits only if it expects a net gain.

A simplified game between the taxpayer and the Tax Administration can be designed (Graetz et al., 1986) assuming that there are only two income levels, W_H and W_L. Moreover, only a fraction p of taxpayers play strategically, cheating whenever it pays, while the other ones always comply. This feature seems in line with empirical findings. While models of choice under uncertainty would suggest that, with actual enforcement parameters, every taxpayer should be a partial evader (unless he has an infinite risk aversion), it actually seems as though only a fraction of taxpayers (often much under 50%) do evade to a certain extent.[15]

Let us assume that taxation is lump-sum, and amounts to either T_H or T_L, according to whether reported income was high or low, with $T_H > T_L$, $T_H < W_H$ and $T_L < W_L$. Poor taxpayers are not interested in cheating (as they can be immediately found out, provided that income cannot drop under W_L), while the rich may cheat, mimicking the poor's income report. Population is normalized to 1 and q is the proportion of the rich. If a rich person is audited, her true income is found out with certainty and she must pay the due tax and a fixed fine θ; it is assumed that these liabilities never exceed her true income. Strategic taxpayers, on whom we will focus from now on, have all the same utility function and are risk-averse.

Rich taxpayers cheat whenever the expected utility from reporting the low income is higher than the utility available with certainty if they comply:

$$(1 - \beta)U(W_H - T_L) + \beta U(W_H - T_H - \theta) > U(W_H - T_H) \tag{4}$$

where β is the probability of audit and $U(W)$ is the taxpayer's utility function. Thus, there is a critical value β^* that solves (4) with equality, i.e., at this critical probability β^* a rich taxpayer is indifferent about cheating or not, If actual audit probability is lower, she always cheats; if it is higher, she never cheats.

The Tax Administration bears a constant cost c for every audit, and has no budget constraint. Marginal net benefits of auditing are positive if:

$$\frac{pq\alpha}{(pq\alpha + 1 - q)}(T_H + \theta - T_L) - c > 0 \tag{5}$$

where the fraction refers to the conditional probability that a taxpayer has high income provided that he reported low income (p is the proportion of strategic taxpayers, q that of rich, and α the probability of cheating). In this case too there is a critical value α^* that solves (5) with equality; at α^* the Tax Administration is indifferent about auditing or not. If the probability is higher it always pays auditing, while if it is lower it never pays. Let us consider two possible Nash equilibria:

1. If $\alpha^* > 1$ the conditional probability of finding a rich liar (which is at most 1) could never be high enough to render auditing worthwhile. Thus the Tax Administration does not audit, and rich people respond to the zero audit probability by always cheating. This case occurs, for example, when c is very high, or the proportion of either rich or strategic taxpayers is low.
2. If $\alpha^* < 1$, taxpayers cheat with probability α^*, while the Tax Administration audits with probability β^*. Equilibrium occurs because at these critical values *both* parties are indifferent about the two actions they could adopt. When taxpayers play "evasion" with probability α^*, and the Tax Administration plays "auditing" with probability β^*, neither party has anything to gain from modifying its behavior.

This model thus implies that, when case 2 occurs, audits sometimes discover cheating taxpayers, in accordance with empirical findings. However a less realistic feature of the resulting equilibrium is the indifference of both parties toward tax evasion. It seems more likely that, in practice, evaders believe that they are likely to gain from tax evasion, while the Tax Administration feels that auditing helps to increase revenue.

Within the model considered it is possible to examine the consequences of variation of taxation and enforcement parameters. It is readily checked that an

increase of θ reduces both α^* and β^*, as on the one hand a lower probability of cheating is needed to render the Tax Administration indifferent about auditing or not (as now each audit is more fruitful), while on the other hand a lower probability of audit is enough to render the rich taxpayer indifferent about compliance and evasion (as the taxpayer's utility in case of audit diminishes). This implication can be seen as an extension of the results of the classical models of evasion as a choice under uncertainty (see Section 2) that foresees a reduction of audits as a reaction of the Tax Administration to the increased taxpayer compliance stemming from higher penalties.

As in Section 2, an increase of the tax rate T_H that regards potential evaders (the rich) while T_L remains the same, has ambiguous effects owing to the opposing influences of substitution and income effect upon the taxpayer's choice. Hence the audit probability β^* needed to preserve taxpayer indifference [see equation (4)] could either be higher or lower than before. On the other hand equilibrium probability of evasion α^* diminishes, as the effect of T_H is identical to that of θ in equation (5).

Within this model it is also possible to study the consequences of an increase of the audit cost c, a parameter not considered within the standard approach of Section 2. An increase in c, as intuition suggests, commands a higher equilibrium probability of evasion α^* [see equation (5)], while the equilibrium probability of audit faced by each strategic taxpayer stays the same [see equation (4), where c does not appear].

The game theory approach has been developed in a large stream of literature, which, while still focusing on proportional taxation, has assumed continuous income distributions and introduced various alternative assumptions, about taxpayers risk neutrality or risk aversion, presence or absence of non-opportunistic taxpayers, availability of unlimited or limited budget for the Tax Administration, etc. Some regressive bias in taxation is found (Reinganum and Wilde, 1986). The reason is that, with a continuum of risk neutral strategic taxpayers, the audit probability must decrease (or at most stay the same) if reported income increases, in order to encourage reporting. But if the Tax Administration must remain indifferent about whether or not to audit high income taxpayers with probabilities that decrease with income, the absolute amount of underreporting must also decrease with income, otherwise it will be worthwhile to search high reports more thoroughly. Poor taxpayers thus evade a larger proportion of their income, and contribute to revenue in terms of costly fines. To overcome the regressive bias, the Tax Administration can resort to income indicators to form audit classes. Also policies already considered within the principal agent approach, like the cutoff policy, could be adopted (Landsberger et al., 2000).

The game theory approach, like the principal-agent one, may also involve some problems of consistency in the Tax Administration choices, whenever the

Tax Administration can infer the taxpayer's income from her report. In models which, like the one presented in this section (Graetz et al., 1986), consider the presence of non-strategic taxpayers,[16] however, greater realism is attained, as the Tax Administration cannot infer without investigation whether an income report is honest or not.

7.   PENALTY SETTING

In many respects enforcing taxation is like enforcing law in general. In this field a classical principle, stressed by Becker (1968), says that monetary sanctions should be brought to their maximal level, in order to save on enforcement costs. Efficiency in deterrence is reached by lowering the audit probability as much as possible (which involves consuming resources), while actual expected sanctions are increased enough to discourage crime by raising the legal sanction. An implicit assumption of this approach is that enforcement is designed and implemented by a benevolent dictator, otherwise citizens would be willing to impose limits to powers in this field. The maximal penalty approach has been applied to tax evasion (e.g., Kolm (1973)) and leads, so to say, to the prescription of "hanging evaders with zero probability."[17] Monetary penalties are however often bound in practice by the liable agent's ability to pay, which is usually measured by her income or wealth.

The maximal penalty approach has been criticized in many respects (Garoupa, 1997).

In practice, resorting to maximal penalties seems rare. One explanation, suggested again by Becker, is that social benefits may outweigh costs when law violations are few; in this case penalties should not be prohibitive. This argument may be applied to tax evasion. The possibility that tax evasion may in some circumstances be socially beneficial instead of harmful, fostering economic growth, employment etc., has often been alleged (see Section 4).

Another criticism of the maximal penalty approach is based on marginal deterrence: as Stigler (1970) has pointed out, if greater law violations are not punished more severely, they become attractive for those who have committed a minor violation. It has been alleged by Mookherjee and Png (1992), however, that marginal deterrence could sometimes be provided for by varying the audit probability, in order to make the expected fine fit the crime, while still resorting to maximal legal penalties in order to save on resources. An example of this type of expected penalty differentiation is provided for by the cut-off rule, which implies audit probabilities that vary with reported income. To differentiate expected penalties by varying the probabilities, the enforcing agency must however be able to cheaply collect preliminary information about the offence. On this basis it can choose how far to investigate and thus differentiate the probabilities. Income reports have been considered as the typical

information source that may enable the Tax Administration to conduct this screening. Marginal deterrence through penalties graduation can be dispensed with also in principal agent models that design incentives (e.g., differential taxes) that induce from the outset the self-selection of taxpayers' types (see Pestieau et al., 1994). Moreover, even small audit mistakes could in this case become irrelevant, as long as with maximal penalties virtually no one dares to violate the law and the agent's chance of being audited is virtually nil (Baron and Besanko, 1984).

A further criticism against maximal penalties hinges upon jurors' behavior (Andreoni, 1991). Both legal principles (such as the reasonable doubt test) and psychological dynamics imply that the actual conviction probability decreases *ceteris paribus* as the penalty provided for by the law increases. That is, the higher the penalties, the more jurors become concerned about possible mistakes (convicting an innocent party). Thus maximal penalties are not optimal and they could even be counterproductive. As, however, jurors are also more willing to punish offences the greater they are, there is some scope for increasing deterrence by increasing penalties as the severity of crime rises; that is, penalties should fit the crime.

A major criticism to the maximal penalty approach hinges on risk aversion. For risk neutral taxpayers, monetary sanctions are pure transfers, which cannot reduce efficiency. On the contrary, risk averse individuals who expect monetary sanctions suffer a further welfare loss, in addition to the amount the enforcing agency expects to receive. Risk averse offenders would in fact be ready to pay a higher amount than the expected sanction in order to insure against it. With reference to tax evasion this welfare loss has been referred to as the "excess burden" of tax evasion. If penalties are raised, this burden increases for those who find it advantageous to keep on evading.

Criticism against maximal penalties based on risk aversion have been rebutted, however, by pointing out that if the social benefits from the breach of law are very small[18] or negative, and penalties can be set high enough to avoid offences altogether, with maximal penalties no one would suffer any excess burden (Kaplow, 1992). That is, resorting to maximal penalties, the excess burden borne by risk averse taxpayers would be cancelled out, and not simply reduced, as is the case with penalty reductions. Marginal deterrence could in this case also be ensured by relying upon screening devices (such as differential taxes and audit probabilities based on taxpayers' reports (Pestieau et al., 1998).

Risk-aversion could, however, motivate the renunciation of maximal penalties when violations can easily stem from the agent's mistakes, as is possible with tax evasion. While maximal penalties could induce risk averse taxpayers to be careful to avoid mistakes altogether, this could impose too heavy a cost

on them, so that lower penalties and higher probabilities are preferred in order to reduce taxpayers' utility loss from risk-bearing. This observation, put forth by Kaplow (Kaplow, 1992) and developed by Boadway and Sato (Boadway and Sato, 2000), is certainly relevant with reference to the problems that would arise if a system based on maximal penalties were introduced anew. As mistakes often regard the manner of interpreting the law, particularly when the law is new, very large welfare costs are likely to arise as a consequence of a switch to maximal sanctions.

Monetary sanctions are no longer costless transfers if it is possible to reduce the detection probability by expending resources (in expert advice, political patronage, etc.). In this case, as with other types of costly sanctions such as imprisonment, efficiency could require less than maximal levels (Malik, 1990).

Other standard criticisms against maximal penalties, that is the incentive to corrupt the enforcer or to submit oneself to enforcer's extortions, seem to be mainly examples of the already mentioned problem of marginal deterrence (Mookherjee and Png, 1994).

In order to secure law enforcement, one could also resort to the opposite approach, that is, offering maximal rewards with small probabilities. However, such a promise is not likely to be credible, as it might involve using greater resources than those available to the enforcer (Border and Sobel, 1987). Thus, to maintain credibility, bonuses for compliant taxpayers must be limited, and assume, for example, the form of tax rebates. Rewards to audited honest taxpayers have been shown to be part of an optimal tax and enforcement policy when taxpayers are risk averse (Mookherjee and Png, 1994).

8.   MORAL PRINCIPLES AND SOCIAL FACTORS

In the literature that treats tax evasion a distinction is often made between opportunistic and non-opportunistic taxpayers: the latter always comply, disregarding standard economic calculus. This approach is in line with empirical evidence,[19] which shows that a large proportion of taxpayers are willing to comply. Theory has mainly focused, however, upon opportunistic behavior, and a convincing explanation for why full compliance so often arises as a rule of conduct is still to be proposed. One relevant determinant of compliance is probably represented by moral principles, as long as social factors, which provide a kind of "internal enforcement" to be distinguished from the external one provided for by the parameters considered in the previous sections. The main difficulty seems to reside in the many facets that moral, social and psychological factors assume, so that there is no obvious approach to modelling their contribution, even if their importance is almost unanimously recognized. For example, a taxpayer may feel morally entitled to evading taxes because he feels that retributive justice is violated (he does not receive from the state as much

as he pays), or because distributive justice is violated (the tax burden is not
fairly distributed, either out of legal provisions or out of actual tax payments),
or because justice is violated in tax enforcement (with reference to audit rates,
sanctions, amnesties, etc.). These moral judgements are doubtlessly subject to
interplay with other factors in dictating conduct. One may consider for ex-
ample psychological factors (such as feelings of guilt, shame etc. in case of
tax evasion), selfish calculations (risks involved in evading taxes), other peo-
ple's reactions, etc. The modalities chosen in the literature to model the inter-
play among these factors are varied: some examples will be provided below.
Perhaps the most widely shared assumption pertains to reciprocation, which
means that taxpayers guided by moral values are upset by other people's eva-
sion, and react to it by reducing their compliance in some way.

From a psychological point of view, tax morale is seen as a factor that gives
rise to a sense of guilt when evasion is implemented, and to feelings of shame
in the case of detected evasion. On this basis, Erard and Feinstein (1994b)
amend the standard utility model by introducing weights that reduce the tax-
payer's utility in the state of the world in which he is not caught (taking guilt
into account) and in which he is caught (considering shame). As both weights
increase with the evaded tax and negatively affect utility, this model *ceteris
paribus* predicts a higher compliance than the standard one. The contribution
of each psychological factor varies with the probability of occurrence of the
relevant state of the world: e.g., the higher the probability of being detected,
the heavier the role of shame in reducing the evader's expected utility.

While a psychological approach underlines the internal consequences of
tax evasion, models that focus upon social stigma consider also the role of
external constraints, other than the law enforcement. Social disapproval, that
hits detected tax evaders, is deemed to be larger the larger the number of honest
agents in the society. Interdependences in behavior that arise in this case give
rise to multiple possible social equilibria (Kim, 2003).

The role of moral principles in explaining tax evasion has been examined by
Bordignon (1993) by specifying the fairness rule followed by agents. Each tax-
payer determines a reference proportional income tax, according to the moral
principle of paying as much as she would wish other individuals to pay. This
"Kantian tax" depends on the amounts of public goods to be financed, on tech-
nology and on resources constraint. The Kantian tax is then corrected in order
to reciprocate against deviations of other taxpayers from their Kantian due
payments, which can arise either out of tax evasion or because the actual pro-
portional tax rate does not distribute the tax burden according to the Kantian
rule. The tax perceived as fair is thus the Kantian one corrected for reciprocity
considerations. To decide upon her tax payment, the taxpayer also solves a
standard expected utility problem; then she pays the higher of the two taxes:
the one that maximizes her utility or the fair one. That is, moral principles

imply that the amount of tax paid can never drop below the fair one, even if evasion would be utility enhancing. On the other hand, considerations concerning the riskiness of tax evasion may imply that the evasion which is considered permissible in moral terms is not implemented when it would be utility reducing. Examination of Nash equilibria for this model is complicated, because according to the values of the parameters the agents' choices could be dictated either by morale or by risk considerations. Under given conditions the model indicates that tax evasion should increase with tax rates and arise out of overprovision of public goods.

Pommerehne et al. (1994), in an evolutionary political economy model, assume that morale induces compliance of the (possibly large) share of taxpayers who are intrinsically cooperative or "Good," while there is also a share of "Bad" members of society who are evasion-prone. It is also assumed that good taxpayers reciprocate, choosing evasion as a means of retaliating against other people's non-cooperative behavior. Evolution starts at the constitutional stage, in which rights and duties are agreed upon, and tax evasion is not foreseen. Tax evasion by the "Bad," however, becomes apparent when, in the post-constitutional stage, the amount of public goods falls short of the predetermined one. As a consequence, the fiscal system could collapse, producing mass evasion and no revenue, if a remedy to the frustration of the cooperative attitudes of the "Good" is not quickly introduced. Institutions may play an important role in shaping reactions to tax evasion. While in the Pommerehne et al. (1994) model auditing policies are chosen and implemented by the government within its administrative responsibilities, penalties for tax evasion are seen as a policy issue, submitted to the choice of the median voter. A "Good" decisive voter would thus demand higher penalties for tax evasion the higher is the loss she suffers because of it. If institutions are able to promptly adjust penalties, tax evasion can be recouped at low cost. However, in a representative democracy, the will of the median voter is not perfectly translated into public choice, and this fact could favor the erosion of cooperative attitudes. Instead, according to Pommerehne et al. (1994), direct democracy is better able to react to tax evasion, reaching stable tax revenues, with lower audit rates and higher penalties than representative democracies.

Moral considerations are also relevant for explaining the repentance of evaders, which can result in voluntary evasion disclosure or amnesty participation. Malik and Schwab (1991) model this case by assuming that the taxpayer, while filling in his tax return, is uncertain about the parameters of his utility function, as he is not able to fully evaluate the social and personal consequences of tax evasion. When uncertainty is resolved, even if the taxpayer has not been audited, he may prefer nevertheless to modify his choice, particularly if a tax amnesty reduces the cost of repentance.

The role of moral factors in tax compliance suggests policies against tax evasion aimed at fostering citizens' inner motivations for compliance. Fairness in the exchange relationship between the citizens and the state, in the tax burden distribution and in tax implementation, achieved thanks to the citizen's direct political participation, would thus be the best cure for tax evasion. While these goals can in principle be pursued alongside deterrence against the "Bad," a crowding-out effect may arise if the "Good" perceive interventions aimed at deterrence as external impositions, thus impairing their feelings of voluntary and free contribution to a just cause and their pride for spontaneous good conduct (Frey, 1997).

9.   KINKED INDIFFERENCE CURVES

Within the standard tax evasion model, justification of the great extent of tax compliance often observed in practice (see Section 8), implies the assumption of very high and unrealistic risk aversion coefficients. Tax evasion is not, however, the only field in which the expected utility approach gives rise to this type of problem. Similar shortcomings also arise in other fields where agents decide under risk. Much attention has been devoted to financial markets, where long term analysis of equity prices based on the expected utility approach shows that prices have incorporated very high risk premia, which could only be explained by assuming very high risk aversion coefficients (Cecchetti and Mark, 1990). As it is not easy to reconcile empirical evidence with the standard expected utility approach, and since experimental results often disconfirm it (see Camerer and Ho, 1994), a promising approach seems to be modification of the description of behavior under risk (Epstein, 1992). Without going into the many facets of this stream of research, only a few issues strictly pertaining to tax evasion will be considered here.

Expected utility with rank dependent probabilities (EURDP) models assume that the agent arranges the outcomes of a lottery in increasing order of preference, and perceives the probability of their occurrence deformed according to their rank. Actual probabilities are weighted by the agent to form her perception, with weights that imply an overestimation of small probabilities of facing disappointing outcomes. Specifically, in the field of taxation, EURDP produces an overestimation of the probability of being detected (which is the worst outcome in the tax evasion lottery), which may motivate greater compliance. Figure 6 shows a modified probability function (which refers to the unfavorable outcome), whose parameters have been estimated by Camerer and Ho (1994) on the basis of large experimental data; it exhibits overestimation for small probabilities.

The EURDP approach implies that the taxpayers' indifference curves have a kink on the certainty line.[20] Thus, taxpayers with different budget lines bunch

on the certainty line (Figure 7). To understand why the kink arises, let us consider in Figure 7 the contingent claims budget set *AB* representing the available combinations of the two outcomes provided by a gamble. On the abscissa of each point of the budget set one can read the income *X* the agent receives if the state of the world α occurs, while on the ordinate the income *Y* that she receives if state β occurs. If the point at which the agent is located is below the 45° degree line the more favorable outcome (and thus the one ranked higher) occurs in state α, in which the agent receives income *X,* while above the 45° degree line it occurs in state β, in which she receives income *Y.* For example if the gamble is tax evasion, let us assume that state α is "going undetected," while state β is detection of evasion, as seen in Section 2. This type of gamble is actually played only in the lower part of Figure 7 (that is, on or under the 45° degree line), as the taxpayer either evades (choosing a point under the 45° degree line) or complies (choosing a point on the 45° degree line), while he has no advantage in paying more taxes than those owed. For an evader, income *X*, which he enjoys if detection does not occur, is clearly higher than *Y,* and thus it is ranked higher. If overcompliance would however
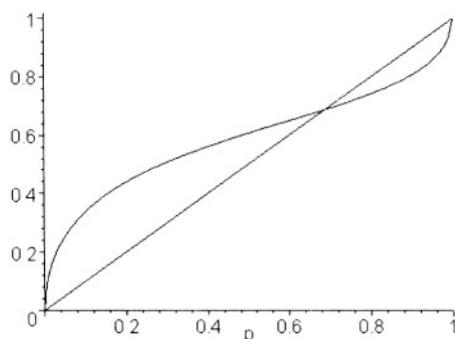
FIGURE 6. Perceived Probability.
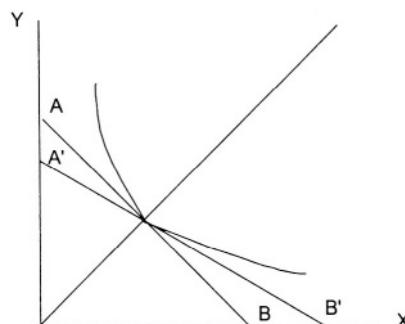
FIGURE 7. Kinked Indifference Curve.
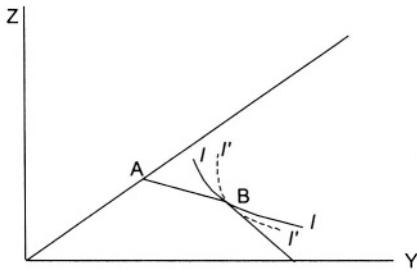
FIGURE 8. Kinked Budget.

command a reward, it would be worthwhile to consider a point in the upper part of the diagram (above the 45° degree line). For an overcomplier, going undetected would be a disappointing outcome, and she would rank Y higher than X. As the rank of outcomes changes as we move from below to above the 45° degree line, the probability weighting of X and Y is reversed, and this fact produces a kink in the indifference curve on the 45° degree line. In fact two expected utility functions are considered, one that weighs the probability of X as though it is a good outcome and that of Y as though it is a bad one; the other one reverses the ranking and thus the weights.

With reference to tax evasion, overweighting given to the probability of audit within the EURDP lowers the risk aversion needed to generate small levels of tax evasion; full compliance could also arise if some role is played by moral values (Bernasconi, 1998).

In the analysis developed thus far, kinks in the indifference curves play a key role. Bunching effects could however stem instead from tax schedules. Progressive taxation implies a kinked budget set. Thus, even taxpayers with different tax attitudes can bunch at a kink; the more progressive the tax, the larger should be this bunching effect that prevents even more risk averse taxpayers (with flatter indifference curves) from choosing a higher compliance than the less risk averse ones (Figure 8).

10.   OPTIMAL TAXATION

The classical optimal taxation problem assumes that a benevolent dictator aims at maximizing the value of a social welfare function, and resorts to taxes in order to finance public goods and income redistribution. However lump-sum efficient taxation on potential income is impossible due to an information asymmetry: only realized income can be observed by the dictator, and taxpayers may hide their potential income by choosing low work effort in order to reduce their taxable income or to qualify for net transfers. Hence taxation, either on income or goods, originates an efficiency loss (excess burden) and an equity-efficiency trade-off follows.

The optimal taxation problem has been reformulated in order to take into account another type of asymmetry. Only taxpayers know their realized income (which in this case is assumed to be exogenous) while the dictator receives a report, which can be false. The dictator can choose tax rates, penalties and auditing policies in order to maximize the value of a social welfare function. As auditing is costly, whenever audits cannot be approximately zero by relying on deterrence produced by maximal penalties, an efficiency cost must be born. There is again an equity-efficiency trade-off, as higher and more progressive taxes are likely to encourage cheating and thus to increase audit costs, which represent an excess burden.

Let us focus first of all on efficiency in taxation, leaving equity aside. A tax system can be termed audit efficient (Chander and Wilde, 1998) if no other system provides at least the same gross revenue, while lowering audit probabilities (with a strict reduction for at least one taxpayer). With risk neutral taxpayers and penalties limited by the taxpayer's income, efficient schemes imply a (weakly) concave tax function (Chander and Wilde, 1998). Intuition about this result relies on the fact that on the one hand penalties, when set at their maximal level, cannot rise faster than income. Moreover, the audit probability must stay the same or decrease as reported income increases, in order to encourage disclosure (see Section 5). Let us consider the incentive for a taxpayer to reveal her true income Y instead of reporting a lower income X:

$$R(Y) \leq (1 - p(X))T(X) + p(X)Y \tag{6}$$

where $R(Y)$ is total payment made if true income is reported. On the right hand side of equation (6) there is the expected payment if X is reported; $T(X)$ is the total tax paid if X is reported, $p(X)$ is the audit probability conditional on reported income X, and Y is the payment if detection occurs, set at its maximal feasible level. Inequality (6) says that a risk neutral taxpayer, to be willing to report her true income, must pay less than the expected payment she should make if she reports X. By rearranging (6) we get:

$$\frac{R(Y) - T(X)}{Y - T(X)} \leq p(X). \tag{7}$$

As $p(X)$ must stay the same or diminish as reported income rises, the marginal payment rate [on the left hand side of (7)]cannot rise; taxation cannot be progressive. Moreover, the marginal payment rate determines the audit probability (Chander and Wilde, 1998). The cut-off rule is an example of an efficient tax scheme (Section 5); when marginal taxation rate is 1 for incomes under the threshold, it commands an audit probability of 1, while over the threshold both the marginal taxation rate and the audit probability drop to 0.

Net revenue maximization (which implies audit efficiency) leads however to progressive taxation in specific cases (Hindriks, 1999). If taxpayers have relative risk aversion increasing with income, it becomes easier to deter them at high income levels. A similar effect arises with auditing technique which, while being imperfect, are more effective as the evaded amount increases.

So far tax evasion has been considered as a factor that limits feasible redistribution. One may wonder, however, whether tax evasion might instead perform any useful role. If taxpayers differ in their risk attitudes, screening them accordingly might help in revealing their contributive capacity. Brito et al. (1995) demonstrate that randomization in tax schedules may be Pareto improving, as it can help to induce the self-selection of taxpayer types on the basis of their different attitudes toward risk. They refer however to random tax schedules, assigned before each taxpayer chooses her effort level, while randomness introduced by tax evasion occurs afterwards and offers thus a narrower set of choices. Pestieau et al. (1994) suggest the introduction in the tax code of random taxes (whose amount would be established by running lotteries), in order to secure screening benefits of randomization, while reducing the audit costs involved in implementing the tax evasion lottery.

A large body of literature has studied the optimal taxation problem assuming that available tax instruments are limited to the linear income tax, in models where income is endogenous (e.g., Sandmo (1981), Cremer and Gahvari (1996)), and equity considerations are introduced. When tax evasion can arise, for example because penalties are bounded, the equity problem becomes more complex. Should the social welfare function include tax evaders as well as compliers? What weight should be given to their utility? Should evaluation be made assuming an *ex-ante* or an *ex-post* perspective (before or after audits have been accomplished)? With reference to efficiency considerations, when income is endogenous, one must remember that tax evasion allows reactions to taxation that may partially offset the classic excess burden due to reduced work effort (for example by enlarging labor supply in the black market). In fact results about the desirable degree of tax progressivity are not clear cut.

11.   POLITICAL AND CONSTITUTIONAL IMPLICATIONS

Taxation gives rise to relevant political problems and is interwined with politics at large. Historically, the politicization of fiscal problems assumed even extreme forms; there were fiscal protests at the very beginning both of the French Revolution and of the American War for Independence. Less impressive but still visible cases of collective refusal of a tax can also be found nowadays. Besley et al. (1997) analyze the wide noncompliance with the poll tax introduced in 1989 in the United Kingdom, and find evidence of the role of

political factors (such as Conservative vote and Conservative control of local authority), as well as of social influences (proximity of high evasion districts) in explaining taxpayer behavior.

Buchanan and Faith (1987) include tax evasion among the forms of internal exit-option that citizens can resort to in order to resist exploitation perpetrated by a monopolistic government, ruled by a coalition that provides public goods but appropriates the surplus of the public budget. Like voting-by-the feet, internal exit is potentially capable of eliminating exploitation and inducing efficiency, whereas many factors may prevent the attainment of such a result. Tax evasion, unlike secession, which can prelude the foundation of a new state, cannot *per se* give rise to a new political entity (unless we focus on illegal markets as systems characterized by both tax evasion and the supply of some substitutive collective good); thus it can have a role only in determining the size and the composition of the ruling coalition, and in constraining feasible taxation. The effectiveness of internal exit threats depends on various economic factors. Resort to tax evasion is likely to be more effective when perpetrated by a rich person, who has more income to subtract from taxation, or by a person who has a demand for public goods lower than members of the ruling coalition, as this fact would command a lower tax rate.

A beneficial role of tax evasion in disciplining a Leviathan government can be pointed out specifically for indirect taxation. Let us consider for example an excise tax on a good characterized by rigid demand. If tax rate increases trigger increasing tax evasion, eventually tax revenue will shrink. This reaction mimics that of elastic demand, which is suited to resisting tax exploitation.[21]

Assuming a Leviathan government, tax evasion may thus produce beneficial effects, by reducing exploitation and increasing citizens' welfare. Its role instead appears potentially detrimental to equity (and sometimes to efficiency as well) within models that describe political markets as imperfectly competitive, even if non monopolistic. Let us consider for example a probabilistic voting model, in which parties choose policies to maximize their expected number of votes in elections. Focusing on the revenue side of the budget, Hettich and Winer (1997) find that marginal conditions for political support maximization imply equalization of marginal losses of consent from raising an additional dollar of tax revenue among citizens. Opportunities and costs of evading or avoiding taxes are among the factors that influence the citizens response to a given political platform. As these opportunities are different, the tax structure should be differentiated accordingly; tax discrimination would thus ensue. Administrative costs, however, pose a limit to the amount of feasible discrimination, which then assumes the form of special provisions, tax shelters etc., within a given tax structure.

Politicians seeking reelection could also benefit by discriminating expected tax rates (rather than legal ones), by directing tax-enforcing agencies to dis-

criminate their audit policies among social groups, geographical areas etc. in order to maximize political support (Hunter and Nelson, 1995). On the other hand, rent-seeking activities could be undertaken by citizens or interest groups, by expending resources in order to secure lenient audit policies and small penalties for tax evasion. Baumol (1990) indicates tax evasion and avoidance among the prime factors that divert entrepreneurial effort from productive to unproductive activities.

While rent-seeking in tax enforcement is thus likely to foster discrimination in either the legal or the actual tax system, constitutional clauses that prohibit it and impose generality in tax norms may prevent rent-seeking from the outset by destroying its prospective benefits (Buchanan and Congleton, 1998).

Tax Administration and enforcement are bureaucratic activities, and can give rise to typical problems of inefficiency or excess supply. Moreover, tax evasion may be linked to corruption of public officials or be the basis for extortion on their part, etc.[22] Concealing tax evasion activities can also be joined in production with other illegal form of rent protection, such as market entry barriers, cartel formation etc., which also imply efficiency losses (Paul and White, 1994). The extreme case is represented by organized crime, which can provide some kind of social order alternative to the legal one, arbitrating controversies, using threat and violence to enforce agreements otherwise not permissible, and also levying the equivalent of taxes.[23]

12.   EVASION OF INDIRECT TAXES

In this Section a partial equilibrium approach is adopted, focusing on the market where the tax operates. Evasion of indirect taxes may assume the form of smuggling. Some products are sold in the streets, without complying with the tax law, at prices lower than those charged in legal transactions. Illegal markets are in this case characterized by a kind of network externality, that is, the higher the number of people involved in illegal transactions, the lower the risk of each transaction (McLaren, 1998). If enforcement is performed through a fixed number of inspectors who simply control the streets and apply a given monetary sanction to the vendor whenever they observe an illegal deal, then the higher the number of these deals, the lower the probability of being fined. Let us assume that both the illegal and the legal market are competitive; each dealer buys one unit of the good in the international market, where supply is perfectly elastic, and sells it, whether legally or not. In the latter case, expected sanctions are a cost for the vendor. Consequently, when the illegal market grows large and these expected costs fall, we would expect the price to fall, further encouraging the growth of the illegal market. In fact, there might be a sharp growth of illegal markets whenever a given threshold or "critical mass" of transactions is overcome.

The functioning of the market with a high tax t charged per unit of goods sold is depicted in Figure 9. The gross price is reported on the abscissa, and the net price on the ordinate. Gross price implies either the price charged on the legal market (net price plus tax) or the "street price" charged on the illegal market, which comprises the expected liability costs for the vendor. Market equilibria can only lie below the 45° degree line, as the gross price can never be lower than the net one. In the example shown in Figure 9 the (short) straight line that parallels the 45° degree line indicates the net price necessary to keep the legal market in equilibrium. As the tax in Figure 9 amounts to 8.5, the net price is zero when the gross one is 8.5. The net price that keeps the illegal market in equilibrium and renders it viable as a substitute of the legal market must lie between the two parallel lines, since illegal traders can operate only at prices equal to or lower than the legal one.

As the illegal market is competitive, it will be in equilibrium at 0 profits. Assuming that, whenever detection occurs, the vendor's profits are confiscated and a fine is applied, the expected profits of an illegal trader are:

$$E(\pi) = (1 - e(p_g))(p_g - p_n) - e(p_g)\theta$$

where e is the probability of detection, p_g and p_n are the gross and the net price respectively, and θ is the fixed fine. The probability of detection is a function of p_g, with $e' > 0$, that is, the probability is high when the gross price is high (as there are comparatively few transactions), and low when the price is low.

Function $p_n^*(p_g, \overline{\theta})$ that solves $E(\pi) = 0$, and thus keeps the illegal market in equilibrium, is shown in Figure 9. Let us assume that the net price of the good is p_i. The illegal market becomes viable and attracts all the customers at the point where p_i intersects $p_n^*(p_g, \overline{\theta})$, on the right hand side of Figure 9. However, to reach this point, a low enough gross price (which involves a critical mass of illegal transactions) must be applied. To enlarge the illegal market
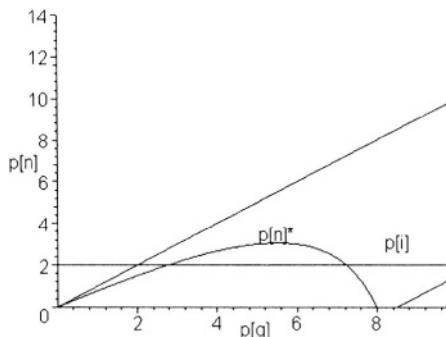
FIGURE 9. Smuggling with a High Tax.

this way, however, the vendors should cooperate, to create the necessary network externality. When the critical mass is reached, also a lower gross price (corresponding to the intersection between p_i and p_n^* $(p_g, \overline{\theta})$ nearest to the origin) becomes viable, as vendors can break-even thanks to further reductions in the probability of detection originated by the large number of transactions.

If the indirect tax is reduced enough (while keeping the sanction and the enforcement effort constant), the multiplicity of equilibria disappears (Figure 10); eventually, the legal gross price may become equal to or lower than the street price, so that legality becomes viable. Thus, within this model, characterized by either full evasion or full compliance, it is always possible to eliminate smuggling altogether by lowering the tax. Moreover, by fixing a tax equal to the expected liability of vendors, there would be no revenue loss for the state. Within an optimal taxation perspective the government should thus always design indirect taxation in order to induce full compliance.

When smuggling is widespread, enforcement turns out to have increasing returns to scale, as it helps to raise prices, thus thinning the illegal market and reducing the number of transactions to police. To exploit scale economies, taxation should thus be concentrated on a few cash-cow goods. Excess burden considerations, however, suggest the opposite approach. As enforcement costs are often very high in developing countries, economies of scale become a decisive factor, so that taxes are often levied just on a few goods. A similar pattern was also widespread in developed countries in the past.

One strong implication of the model under consideration (McLaren, 1998) is that only full evasion or full compliance can arise. Partial compliance could however arise if one considers productive firms that are involved in many transactions. In this case the visibility of each firm is likely to increase with output. By assuming that evasion also involves concealment costs that increase with the firm output, Virmani (1989) shows[24] that partial compliance may arise.
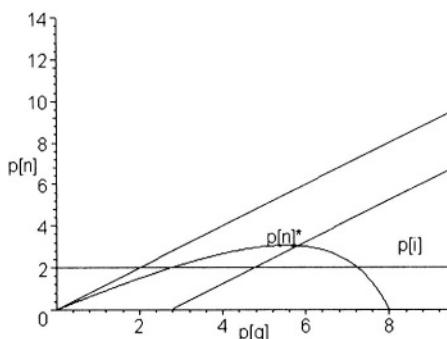
FIGURE 10. Smuggling with a Low Tax.

Partial compliance also arises by considering indirect taxation under monopoly. Marrelli (1984) considers an indirect proportional ad *valorem tax* charged on the price paid by the consumer, and assumes that the monopolist is risk averse; this model gives some support to the idea that a larger output size implies a larger share of reported tax base, as is often considered to be the case in practice. From a policy point of view, a rule that reduces the probability of auditing as declared amounts increase should encourage both output enlargement and compliance. Comparing the rates of declaration for a direct profit tax and for the indirect one that would give the same yield in case of full compliance, Marrelli (1984) shows that the latter is higher than the former under DARA.[25] In fact, since the indirect tax is distortionary, it reduces the monopolist income more than the direct tax,[26] and thus, if risk aversion decreases with income, it induces larger compliance.

13.   TAX AVOIDANCE

Tax avoidance is the exploitation of loopholes and tax shelters provided by the tax law in order to reduce tax liability. This practice can be considered as either somewhat risky (since the enforcing agency could disconfirm the advantage and apply a penalty), or as sure, if it only implies expending resources in order to find the loophole and to exploit it. In the latter case the taxpayer will prefer tax avoidance to compliance only as long as the former gives positive net benefits, which depend on the one hand on the amount of tax that can be avoided, and on the other on administrative and transaction costs. As a high tax liability increases the benefit at a given cost, richer people who must pay high taxes would engage in tax avoiding activities to a larger extent than the poor.[27]

When avoidance and evasion are jointly considered, the standard portfolio choice problem of the taxpayer (see Section 2) becomes more complicated in the presence of tax avoidance opportunities (Alm at al., 1990), as the taxpayer could engage in two activities which could give a positive net return (tax evasion, which is risky, and tax avoidance, which can be modelled as either sure or risky) and in a zero-return sure activity (tax compliance). This setting highly complicates the taxpayer's choice problem[28] and opens the way to ambiguous predictions. For the case in which tax avoidance is not risky Alm (1988) shows that the taxpayer's choice is separable into two stages. At the first stage the taxpayer decides how much to evade. At the second, she allocates the residual income between compliance and avoidance. Comparative static analysis can be applied to examine tax avoidance reactions to enforcement parameter changes that modify tax evasion. For example a reduction of tax evasion due to a higher penalty increases the income available for both compliance and tax avoidance. The latter should increase to the extent in which its net benefits remain positive, thus performing the role of a substitute of tax evasion.

The consequences of tax avoidance on the public budget are similar to those of tax evasion. The attempt at reducing tax avoidance on the part of the state may involve expending resources in audits and investigations. Thus there is also an efficiency loss stemming from tax avoidance, which depends on the resources spent both by the public budget to fight it and by the taxpayers to secure it. The latter may assume the form of political pressure aimed at securing preferential fiscal treatment, tax expenditures, etc., for specific citizens groups (see Section 11).

Among private costs borne to implement tax avoidance, research has focused on third party assistance. Experts may provide information, help in return preparation, represent the taxpayer before the Tax Administration and the courts etc. Also, under third party assistance, tax avoidance may be seen either as a substitute or a complement of tax evasion. Reinganum and Wilde (1991) note that raw evidence suggests a link between taxpayer use of preparers or practitioners and noncompliance. They report U.S. data that show higher noncompliance, both as a percentage of number of reports and as a percentage of the amount of the reported tax, for taxpayers who resorted to third-party assistance. Erard (1997), using a 1982 IRS TCMP data file for the U.S.A., estimates higher success by examiners in detecting deliberate evasion on self-prepared returns than on paid-prepared returns, thus confirming the effectiveness of third party assistance in case of tax evasion. Statistical analysis has shown that among the factors that explain reliance on third party assistance, there are typical indicators of possible tax avoidance, such as the number of exemptions and deductions.[29] The demand for third party assistance is also positively influenced by the taxpayer's tax rate and income (Dubin et al., 1992). Erard (1993) finds however that income level is not significative, while income sources are (business, farm, rental or royalties). Klepper et al. (1991) empirical analysis indicates that tax practitioners assistance favors compliance in cases unambiguously regulated by the law, while it favors non compliance whenever the law is ambiguous.

Erosion of tax bases through tax reductions and exemptions is extensive in developing countries. For example, Gauthier and Gersovitz (1993) list many special provisions available in Cameroon, which even comprise a *convention special* (special convention) under which a businessman may reach any sort of agreement with the Tax Administration about the amount of taxes to be paid, which means that no constraint exists on preferential treatment. Through an analysis based on questionnaires, Gauthier and Gersovitz (1993) find that small business in Cameroon are largely involved in tax evasion, while larger business have the best privileges and conventions and are more likely to resort to tax avoidance. More advantages are enjoyed by firms with public or foreign participation, a fact that seems attributable to the role of political factors in creating tax avoidance opportunities.

The economic meaning of distinguishing tax evasion from tax avoidance vanishes when tax evasion can be coupled with concealment expenditures that in practice entirely eliminate any risk of sanctions. This case is relevant for developing and high evasion countries, where there is a large informal sector made up of small firms, which ordinarily resort to tax evasion and do not comply with regulation. Concealment technology may explain the dualism that characterizes such economic systems, in which only firms which work in the formal sector, complying with the law, reach a sizeable dimension. In the Fortin et al. (1997) general equilibrium model, firms that produce one good using only labor differ in their managerial capacities, so that the more able ones have a higher marginal labor product. Proportional taxes are levied on wages and profits, and a minimum wage is provided for by the law; firms can however escape taxes and regulation by bearing concealment costs eliminating any risk of sanctions. Concealment technology implies costs strictly increasing with the size of the firm (measured by the hired labor). Within this model, under mild conditions, a self selection occurs; less able firms choose to evade and to reach only a small dimension, while more able ones comply and reach larger dimensions. Tax increases (balanced by uniform lump-sum transfers that keep the public budget in equilibrium) enlarge the number of evading firms and negatively affect the GDP, as the marginal product of labor *ceteris paribus* is lower in the informal sector, due to concealment costs. This negative effect must however be sized up considering that the GDP could also be adversely affected by tax increases in the absence of any informal sector, through excess burden effects; thus the comparison must be made between two types of distortions. Through a simulation, Fortin et al. (1997) show that the GDP reduction produced by increased tax evasion can be greater, whereas the presence of an informal sector yields higher employment, thanks to the wage flexibility it provides.

14.   CONCLUSIONS

Since the seminal analysis of Allingham and Sandmo (1972), a large body of literature has focussed on tax evasion. A clearer understanding of this topic, based both on theoretical and empirical research, had a significant influence on the evolution of public finance in the Eighties and later. Attempts at slowing public sector growth and reducing income tax progressivity, sometimes considered as a consequence of the lesson of optimal taxation theory, may well have been inspired, among other factors, by the diffusion in developed as well as developing countries of tax avoidance, tax evasion and black markets. Revenue leakages they originate challenge the possibility for the state to engage in the production of public goods and in income redistribution.[30]

While tax evasion analysis shares many characteristics with the study of more standard types of excess burden originated by taxation, it also has many distinctive characteristics.

Riskiness of tax evasion pairs it with choices under risk in the field of insurance and finance. While many results obtained in the latter fields have been either restated or amended with reference to tax evasion, there is probably still room for further contributions, for example with reference to multiple choice settings, risk prevention expenditures, generalized and non expected utility approaches, and experimental analysis of behavior under risk.

Policies aimed at curbing tax evasion and pursuing government goals have been studied either through principal agent models or game theory. In this field there is some parallelism with studies of other distortions originated by taxation in terms of reduced work supply. A specific feature of tax evasion lies in the large information set that the state can exploit to curb it, resorting to reports, investigations, signals of contributing capacity etc. Other relevant features of tax evasion pertain to information externalities, interdependence of choices among individuals, and the herding behavior that may characterize it. Moreover, the distortions that tax evasion originates interact with other distortions in a complex way, with a net effect on efficiency which is not necessarily negative.

If tax evasion is a game, it is not however a game against nature, but instead a game whose pay-offs depend on public choice. The most specific feature of tax evasion resides in the social, moral and political problems that it involves. It is a form of law violation, often practiced by a sizable share of the population, which questions the coercive power of the state, while sometimes offering positive contributions to GDP and employment. Lenient interventions against tax evasion may in turn be dictated by political exchange and rent-seeking activities.

The reactions to tax evasion are thus necessarily multifold, ranging from the recognition that in specific circumstances it can have beneficial effect, to the design of optimal tax systems and enforcement techniques aimed at reducing its scope, to the strengthening, through constitutional rules, of the basis of legitimacy and consent toward the power to tax, in order to minimize determinants of tax evasion from the outset.

NOTES

1. Literature on tax evasion has been reviewed, among others, by Pyle, 1991; Andreoni et al., 1998; Franzoni, 2000; Slemrod and Yitzhaki, 2000.
2. With equality the taxpayer is indifferent between cheating or not, as her expected payment stays the same in both cases: it is however usually assumed that compliance arises. On this topic see also Section 6.

3. Whose utility function is obtained through a monotonic concave transformation of a utility function exhibiting risk aversion.

4. Provided that hours of work in the two sectors are perfectly substitutable, the decisions concerning the total number of hours to work and their allocation are separable when $\partial^2 (U_2/U_1)/\partial c^2 = 0$, where U_2 is marginal utility of leisure and U_1 is marginal utility of consumption (see Cowell, 1990).

5. If net income is a normal good, the taxpayer will not choose a higher amount when he is impoverished by a higher taxation (while obviously he can increase the gross income amount, raising labour supply).

6. Johnson et al. (1999) assume that it is the productivity in the official sector which is likely to drop, if the amount of public goods is reduced as a consequence of the revenue shortage due to tax evasion. Examples would be provided by some ex-Soviet Union countries, where legal economic activity is negatively affected by insufficient provision of social order.

7. For the implications of assuming a lower productivity in the evadable sector, see Section 13.

8. Because with CRRA absolute risk aversion declines when income increases.

9. There could be in fact a double peaked Laffer curve, as there are two possible switches (from evasion to compliance and from compliance as entrepreneurs to the sure occupation).

10. The implications of assuming instead risk aversion have been discussed by Townsend, 1979 and Mookherjee and Png, 1989.

11. As revenue can still be raised by raising the tax until it is lower than the cut-off revenue: e.g., the taxpayer who reports income i^* is not fully exploited if the tax is lower than i^*.

12. The reciprocal of the hazard rate, $(1 - G(Y))/g(Y)$, where $G(Y)$ is the cumulative distribution function of income Y and $g(Y) = G'(Y)$, must be decreasing in Y.

13. As stated by the sentence "*de minimis non curat praetor.*"

14. A similar argument is put forth by Hindriks et al. (1999) for the case in which a tax inspector could try to extort payment through the threat of overreporting the taxpayer's true income: it is harder to credibly ovverreport incomes the higher they become.

15. On this topic, see Section 8.

16. On this topic see also Erard and Feinstein, 1994a.

17. Whereas hanging is costly! The principle refers in fact to monetary sanctions.

18. Although there are small positive social benefits that may be lost if the violation is eradicated, the elimination of the excess burden may produce a benefit that overcomes this loss.

19. The majority of taxpayers aim at full compliance according to evidence considered in Feinstein, 1991, based on TCPM data for the U.S.A.; moreover, evasion mainly regards small amounts. Evasion is larger in other countries, however, and particularly in the third world (Burgess and Sternn, 1993).

    Experiments confirm that some people do not evade even when detection cannot occur (Alm et al., 1992), while fairness considerations are significative in explaining compliance (Alm et al., 1993).

20. Utility function in this case exhibits the so-called first order risk aversion. The risk premium for a binary gamble is proportional to the standard deviation, rather than the variance of the gamble (Epstein and Zin, 1990).

21. From the alternative perspective of optimal taxation, Cremer and Gahvari (Cremer and Gahvari, 1993) suggest amendments to the Ramsey rule to take tax evasion into account.

22. For a game theory approach to this topic, see Chander and Wilde, 1992 and Hindriks, 1999.

23. See Fiorentini and Peltzman, 1995. Lump-sum taxation would be the preferred form, in order to avoid the costs of monitoring involved by other tax bases. Tax evasion is thus a concern for the mafia too!

24. The tax in this case is a constant fraction of the gross price.

25. The reporting rate considered is always interior (neither zero nor one).
26. For an analysis of assumptions needed for the neutrality of profit taxes with tax evasion in a monopolistic market see Lee, 1998.
27. Note that a similar argument applied to tax evasion would be misleading, because among the tax evasion costs one must also consider risk bearing. Both the amount of risk involved as well as the attitudes toward it may change when income varies, in ways that preclude a general conclusion about the effects upon tax evasion (see Section 2).
28. Cowell (1990) considers that evasion should not ensue when there is avoidance, as the latter reveals that the taxpayer has at least the income level needed to pay avoidance costs and draws the auditor's attention over the report.
29. Increases in local, state and real estate taxes increase the demand for tax practitioners, while they reduce the demand for public assistance services (Dubin et al., 1992).
30. See Slemrod, 1994, who, however, considering that the effects of tax avoidance upon the desirable degree of progressivity are ambiguous, suggests that equity considerations could justify higher enforcement efforts to sustain redistribution.

REFERENCES

Allingham, M. G. and Sandmo, A. (1972). "Income Tax evasion: a Theoretical Analysis." *Journal of Public Economics* 1: 323-338.

Alm, J. (1988). "Compliance Costs and the Tax Avoidance-Tax Evasion Decision." *Public Finance Quarterly* 16: 31-66.

Alm, J., Jackson, B. R., and McKee, M. (1993). "Fiscal Exchange, Collective Decision Institutions, and Tax Compliance." *Journal of Economic Behavior and Organization* 22: 285-303.

Alm, J. and McCallin, N. J. (1990). "Tax Avoidance and Tax Evasion As a Joint Portfolio Choice." *Public Finance* 45: 193-200.

Alm, J., McClelland, G. H., and Schulze, W. D. (1992). "Why Do People Pay Taxes?" *Journal of Public Economics* 48: 21-38.

Andreoni, J. (1991). "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?" *Rand Journal of Economics* 22: 385-395.

Andreoni, J., Erard, B., and Feinstein, J. (1998). "Tax Compliance." *Journal of Economic Literature* 36: 818-860.

Baron, D. P. and Besanko, D. (1984). "Regulation, Asymmetric Information and Auditing." *Rand Journal of Economics* 15: 447-470.

Baumol, W. J. (1990). "Entrepreneurship: Productive, Unproductive, and Destructive." *Journal of Political Economy* 98 (Part 1): 893-921.

Becker, G. S. (1968). "Crime and Punishment: an Economic Approach." *Journal of Political Economy* 76: 169-217.

Bernasconi, M. (1998). "Tax Evasion and Orders of Risk Aversion." *Journal of Public Economics* 67: 123-134.

Besley, T., Preston, I., and Ridge, M. (1997). "Fiscal Anarchy in the UK: Modelling Poll Tax Noncompliance." *Journal of Public Economics* 64: 137-152.

Boadway, R. and Sato, M. (2000). "The Optimality of Punishing Only the Innocent: the Case of Tax Evasion." *International Tax and Public Finance* 7: 641-664.

Border, K. C. and Sobel, J. (1987). "Samurai Accountant: a Theory of Auditing and Plunder." *Review of Economic Studies* 54: 525-540.

Bordignon, M. (1993). "A Fairness Approach to Income Tax Evasion." *Journal of Public Economics* 52: 345-362.

Buchanan, J. M. and Congleton, R. D. (1998). *Politics By Principle, Not Interest: Toward Nondiscriminatory Democracy.* Cambridge: Cambridge University Press.

Buchanan, J. M. and Faith, R. L. (1987). "Secession and the Limits of Taxation: Toward a Theory of Internal Exit." *American Economic Review* 77: 1023-1031.

Burgess, R. and Sternn, N. (1993). "Taxation and Development." *Journal of Economic Literature* 31: 762-830.

Brito, D. L. et al. (1995). "Randomization in Optimal Income Tax Schedules." *Journal of Public Economics* 56: 189-223.

Camerer, C. F. and Ho, T. H. (1994). "Violations of the Betweenness Axiom and Nonlinearity in Probability." *Journal of Risk and Uncertainty* 8: 167-196.

Cecchetti, G. S. and Mark, N. C. (1990). "Evaluating Empirical Tests of Asset Pricing Models: Alternative Interpretations." *American Economic Review Papers and Proceedings* 80: 48-51.

Chander, P. and Wilde, L. (1992). "Corruption in Tax Administration." *Journal of Public Economics* 49: 333-349.

Chander, P. and Wilde, L. L. (1998). "A General Characterization of Optimal Income Tax Enforcement." *Review of Economic Studies* 65: 165-183.

Clotfelter, C. T. (1983). "Tax Evasion and Tax Rates: an Analysis of Individual Returns." *Review of Economics and Statistics* 65: 363-373.

Cowell, F. A. (1985). "Tax Evasion with Labour Income." *Journal of Public Economics* 26: 19-34.

Cowell, F. A. (1990). *Cheating the Government.* Cambridge, MA: MIT Press.

Cowell, F. A. and Gordon, J. P. F. (1995). "Auditing with 'Ghosts.'" In *The economics of organised crime,* Fiorentini, G. and Peltzman, S., eds., pp. 185-196. Cambridge: Cambridge University Press.

Cremer, H. and Gahvari, F. (1993). "Tax Evasion and Optimal Commodity Taxation." *Journal of Public Economics* 50: 261-275.

Cremer, H. and Gahvari, F. (1996). "Tax Evasion and the Optimum General Income Tax." *Journal of Public Economics* 60: 235-249.

Dubin, J. A. et al. (1992). "The Demand for Tax Return Preparation Services." *Review of Economics and Statistics* 74: 75-82.

Epstein, L. G. (1992). "Behaviour under Risk: Recent Developments in Theory and Applications." In *Advances in Economic Theory: Sixth World Congress of The Econometric Society,* Laffont, J. J., ed., Vol. II, pp. 1-63. Cambridge: Cambridge University Press.

Epstein, L. G. and Zin, S. E. (1990). "'First-Order' Risk Aversion and the Equity Premium Puzzle." *Journal of Monetary Economics* 26: 387-407.

Erard, B. (1993). "Taxation with Representation: an Analysis of the Role of Tax Practitioners in Tax Compliance." *Journal of Public Economics* 52: 163-197.

Erard, B. and Feinstein, J. (1994a). "Honesty and Evasion in the Tax Compliance Game." *Rand Journal of Economics 25:* 1-19.

Erard, B. and Feinstein, J. (1994b). "The Role of Moral Sentiments and Audit Perceptions in Tax Compliance." *Public Finance* 49 (Suppl.): 70-89.

Erard, B. (1997). "Self-Selection with Measurement Errors: a Microeconometric Analysis of the Decision to Seek Tax Assistance and its Implications for Tax Compliance." *Journal of Econometrics* 81:319-356.

Feinstein, J. S. (1991). "An Econometric Analysis of Income Tax Evasion and Its Detection." *Rand Journal of Economics* 22: 14-35.

Fiorentini, G. and Peltzman, S., (1995). "The Economics of Organised Crime: Introduction." In *The Economics of Organised Crime,* Fiorentini, G. and Peltzman, S., eds., pp. 1-30. Cambridge: Cambridge University Press.

Fortin, B., Marceau, N., and Savard, L. (1997). "Taxation, Wage Controls and the Informal Sector." *Journal of Public Economics* 66: 293-312.

Franzoni, L. A. (2000). "Tax Evasion." In *Encyclopedia of Law and Economics,* Bouckaert, B. and De Geest, G., eds. Cheltenham, U.K.: Elgar.

Frey, B. (1997). *Not just for the Money: an Economic Theory of Personal Motivation,* Cheltenham, U. K. and Lyme, N. H., eds. Elgar distributed by American International Distribution Corporation Williston.

Garoupa, N. (1997). "The Theory of Optimal Law Enforcement." *Journal of Economic Surveys* 11:267-295.

Gauthier, B. and Gersovitz, M. (1993). "Revenue Erosion through Exemption and Evasion in Cameroon." *Journal of Public Economics* 64: 407-424.

Graetz, M. J., Reinganum, J. F., and Wilde, L. L. (1986). The Tax Compliance Game: toward an Interactive Theory of Law Enforcement." *Journal of Law, Economics and Organization* 2:1-32.

Greenberg, J. (1984). "Avoiding Tax Avoidance: a (Repeated) Game-theoretic Approach." *Journal of Economic Theory* 32: 1-13.

Harrington, W. (1988). "Enforcement Leverage when Penalties are Restricted." *Journal of Public Economics* 37: 29-53.

Hettich, W. and Winer, S. L. (1997). "The Political Economy of Taxation." In *Perspectives on Public Choice: A Handbook,* Mueller, D. C., ed., pp. 481-525. Cambridge: Cambridge University Press.

Hindriks, J. (1999). "On the Incompatibility between Revenue Maximsation and Tax Progressivity." *European Journal of Political Economy* 15: 123-140.

Hindriks, J., Keen, M., and Muthoo, A. (1999). "Corruption, Extortion and Evasion." *Journal of Public Economics* 74: 395-430.

Hunter, W. J. and Nelson, M. (1995). "Tax Enforcement: a Public Choice Perspective." *Public Choice* 82: 53-67.

Johnson, S., Kaufmann, D., and Zoido-Lobatòn, P. (1999). "Corruption, Public Finances and the Unofficial Economy." *Policy Research World Bank WP* n. 2169.

Kaplow, L. (1992). "The Optimal Probability and Magnitude of Fines for Acts that Definitely are Undesirable." *International Review of Law and Economics* 12: 3-11.

Klepper, S., Mazur, M., and Nagin, D. (1991). "Expert Intermediaries and Legal Compliance: the Case of Tax Preparers." *Journal of Law and Economics* 34: 205-229.

Kim, Y. (2003). "Income Distribution and Equilibrium Multiplicity in a Stigma-Based Model of Tax Evasion." *Journal of Public Economics* 87: 1591-1616.

Kolm, S. (1973). "A Note on Optimum Tax Evasion." *Journal of Public Economics* 2: 265-270.

Lacroix, G. and Fortin, B. (1992). "Utility-based Estimation of Labour Supply Functions in the Regular and Irregular Sectors." *Economic Journal* 102: 1407-1422.

Landsberger, M., Monderer, D., and Talmor, I. (2000). "Feasible Net Income Distributions Under Income Tax Evasion: An Equilibrium Analysis." *Journal of Public Economic Theory* 2: 135-153.

Lee, K. (1998). "Tax Evasion, Monopoly, and Nonneutral Profit Taxes." *National Tax Journal* 5: 333-338.

Malik, A., (1990). "Avoidance, Screening and Optimum Enforcement." *Rand Journal of Economics* 21: 341-353.

Malik, A. and Schwab, R. (1991). "The Economics of Tax Amnesties." *Journal of Public Economics* 46: 29-49.

Marrelli, M. (1984). "On Indirect Tax Evasion." *Journal of Public Economics* 25: 181-196.

McLaren, J. (1998). "Black Markets and Optimal Evadable Taxation." *Economic Journal* 108: 665-679.

Mookherjee, D. and Png, I. P. L. (1989). "Optimal Auditing, Insurance, and Redistribution." *Quarterly Journal of Economics* 104: 399-415.

Mookherjee, D. and Png, I. P. L. (1992). "Monitoring vis-a-vis Investigation in Enforcement of Law." *American Economic Review* 82: 556-565.

Mookherjee, D. and Png, I. P. L. (1994). "Marginal Deterrence in Enforcement of Law." *Journal of Political Economy* 102: 1039-1066.

Paul, C. and White, L. (1994). "Illegal Markets and the Social Costs of Rent-Seeking." *Public Choice* 79:105-115.

Pestieau, P. and Possen, U. (1991). "Tax Evasion and Occupational Choice." *Journal of Public Economics* 45: 107-125.

Pestieau, P., Possen, U. M., and Slutsky, S. M. (1994). "Optimal Differential Taxes and Penalties." *Public Finance* 49 (Suppl.): 15-27.

Pestieau, P., Possen, U. M., and Slutsky, S. M. (1998). "The Value of Explicit Randomization in the Tax Code." *Journal of Public Economics* 67: 87-103.

Pommerehne, W. W., Hart, A., and Frey, B. (1994). "Tax Morale, Tax Evasion and the Choice of Policy Instruments in Different Political Systems." *Public Finance* 49 (Suppl.): 52-69.

Pyle, D. J. (1991). "The Economics of Taxpayer Compliance." *Journal of Economic Surveys* 5: 163-198.

Reinganum, J. F. and Wilde, L. L. (1985). "Income Tax Compliance in a Principal-Agent Framework." *Journal of Public Economics* 26: 1-18.

Reinganum, J. F. and Wilde, L. L. (1986). "Equilibrium Verification and Reporting Policies in a Model of Tax Compliance." *International Economic Review* 27: 739-760.

Reinganum, J. F. and Wilde, L. L. (1991). "Equilibrium Enforcement and Compliance in the Presence of Tax Practitioners." *Journal of Law, Economics and Organization* 7: 163-181.

Sandmo, A. (1981). "Income Tax evasion, Labour supply, and the Equity-Efficiency Tradeoff." *Journal of Public Economics* 16: 265-288.

Scotchmer, S. (1987). "Audit Classes and Tax Enforcement Policy." *American Economic Review* 77: 229-233.

Slemrod, J. (1994). "Fixing the Leak in Okun's Bucket: Optimal Tax Progressivity when Avoidance Can Be Controlled." *Journal of Public Economics* 55: 41-51.

Slemrod, J. and Yitzhaki (2000). "Tax Avoidance, Evasion and Administration." *NBER Working Paper* N. 7473.

Stigler, G. J. (1970). "The Optimum Enforcement of Laws." *Journal of Political Economy* 78: 526-536.

Townsend, R. (1979). "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory* 21: 265-293.

Virmani, A. (1989). "Indirect Tax Evasion and Production Efficiency." *Journal of Public Economics* 39: 223-237.

Yaniv, G. (2003). "Auditing Ghosts by Prosperity Signals." *Economic Bulletin* 8: 1-10.

Yitzhaki, S. (1987). "On the Excess Burden of Tax Evasion." *Public Finance Quarterly* 15: 123-137.

Chapter 11

PUBLIC ENTERPRISE: RETROSPECTIVE REVIEW AND PROSPECTIVE THEORY

Scott Hinds
University of North Carolina at Chapel Hill
shinds@email.unc.edu

Nicolas Sanchez
College of the Holy Cross
nsanchez@holycross.edu

David Schap
College of the Holy Cross
dschap@holycross.edu

Abstract   Most of the literature on public enterprise seeks to set forth rules to govern their creation and their pricing. This chapter starts by describing the contours of this literature, but gives the bulk of its attention to exploring the potential contributions that public choice theorizing might make to our understanding of the actual conduct of public enterprises, conduct that often varies from what the normative literature prescribes.

Keywords: Asset seizure, property rights paradigm, public enterprise, public choice theory, quangos

JEL classification: H20, H50

1.   INTRODUCTION

Public enterprise is a phenomenon whose effects are multifaceted and have been considered in a variety of traditional fields of economics, such as Industrial Organization, Labor, Development, and Comparative Systems. In the exceedingly large and diverse literature on the topic, one finds these organizations alternatively referred to as government enterprises, public corporations and state-owned enterprises (or SOEs). Interestingly, V. V. Ramanadham, an

outstanding contributor to the economics literature on the subject of public enterprise (as noted in Heath, 1990), wrote that it is an "inadequately appreciated fact that public enterprise is an aspect of public finance" (1991, p. 1).

"Public enterprise" is in fact a vague term, employed by different people to connote different things at different times. We cull from the existing literature a working definition of the phenomenon, and explore the treatments of its manifestations in the normative (prescriptive) and the positive (descriptive) literatures. The normative approach, sometimes referred to as policy analysis, is considered only briefly in our review, and with some skepticism. The positive approach is relied on much more heavily in our own analysis, appropriate in a volume relying on a constitutional approach to public finance issues.

The structure of the chapter is very straightforward. After defining public enterprise, we survey and critique the earlier, normative work. Following this is an exploration of the property rights approach to a positive theory of public enterprise, as undertaken by (Alchain, 1965), De Alessi (1974, 1977, 1980, 1982) and a host of others during the 1970s and 1980s. We then go on to specify some shortcomings of the property rights approach that were addressed in the literature in the late 1980s and early 1990s. Next we sketch the features of a positive theory of public enterprise grounded in public choice analysis. Our hope in this journey is to demonstrate that public choice theory does indeed provide a useful and unified way to think about the factors and actors that induce the creation of public enterprises and determine the ways in which they are run. We conclude with some brief remarks about recent developments related to public enterprise in various parts of the world.

2.   WHAT IS PUBLIC ENTERPRISE?

Before we can delve into any useful discussion, public enterprise must be explicitly defined. Some authors have been particularly loose with the label: William G. Shepherd goes as far as to say that the exhibition of any degree of "publicness" in an enterprise's costs, controls, ownership or management would classify that enterprise as public (Shepherd, 1966, pp. 36-37). V. V. Ramanadham, too, argues simply that 'public enterprise' refers to "an organization which combines 'publicness' and 'enterprise'" (1991, p. 3). However, Ramanadham (1991, p. 4) departs from Shepherd by claiming that financial viability and a relationship between the costs a public enterprise incurs and the price it charge—in fact the very elements exhibiting 'enterprise'—distinguishes a public enterprise from other public activities like education and justice. Similarly, Aharoni (1986, p. 6) claims that the three distinguishing characteristics of a public enterprise are government ownership, production of goods and services that ultimately are distributed on a fee basis (i.e., actually sold), and sales revenues that have some connection to underlying costs (c.f. Zeckhauser and Horn, 1989, pp. 10-17).

The definitional problem is compounded by the fact that definitions vary, not only from author to author, but from country to country. Data from the Centre European de l'Enterprise Publique (CEEP) presented by Deiter Bös (1989, pp. 424-431) illustrate the problem perfectly. Data from France follow a strict majority-of-shares rule in defining an enterprise as public (i.e., a majority of shares must be held by the government in order for an enterprise to be regarded as public). German data tend to follow the same rule, but include companies like Volkswagenwerk AG, a company *not* majority government owned but nevertheless regarded as a public enterprise. More strict rules are followed for data in Italy, which include only companies from non-agricultural and non-financial sectors where there is both a public and private sector present. Information from the United Kingdom is restricted to 'public corporations' which are mainly nationalized industries.

The definition we choose to utilize operationally herein was first articulated by R. P. Short, who wrote (1984, p. 111): "Public enterprises have two defining characteristics: they are government owned and controlled; and they are engaged in business activities [i.e., sell their output]." This same definition was subsequently employed by (Schmitz, 1996) and is very close to the definition utilized by the U.S. Department of Commerce. [According to Bös (1989, p. 431)], "Government enterprises are defined as '… the activities of government whose operating costs are at least to a substantial extent covered by the sale of goods and services to the public.'") We are partial to Short's definition for its brevity and efficiency. Of all the characteristics one could consider in classifying an *enterprise* as *public,* only those two, namely government ownership and output sold to customers, seem to us to be logically necessary.

Short (1984, p. 111) touches on the main problem in defining public enterprise—the fact that in most countries there is some government control over literally every firm. The problem is made obvious by taking Shepherd, for example, to his own definitional limit: one could argue that any regulation or taxation is tantamount to 'government control,' and so public enterprise becomes an all-inclusive phenomenon.

It is appropriate here to recount an observation made by Schmitz (1996, pp. 5-7) exploring just what public enterprise means to various economies around the world. Using the data collected by Short, Schmitz finds little correlation between the size of a country's public sector (i.e., public enterprise share of GDP) and productivity, but a strong negative correlation between public enterprise share of *manufacturing* output and productivity. Using data from the 1970s, the United States, Denmark and other industrialized countries had public shares of manufacturing at or close to zero. However in developing countries, the share was significantly higher, even running into 60–70% in some cases. Schmitz offers two possible explanations for this trend: that to

begin economic development, government production in manufacturing is necessary, or alternatively, that the public manufacturing enterprises are precisely the cause of these countries' low productivity. Pointing to the economic history of England and the U.S., Schmitz favors the second interpretation, arguing that government production in manufacturing has not been necessary to economic development (1996, p. 4). Clearly then, the roles of public enterprises differ from country to country—even to the point of inhibiting economic growth—further complicating the task of considering and defining the phenomenon.

In limiting our conceptualization of what constitutes a public enterprise to those organizations that (a) are government owned and controlled and (b) sell their output, a wide range of entities are ignored. For instance, any degree of private investment would distinguish the firm as either a mixed or private enterprise, whereas the provision of goods for free (or for the government's own consumption) would make it an agency or a bureau to our way of thinking. Separate from the public enterprise literature, there extensive writings on such subjects as mixed enterprises, regulation, agencies and bureaus, which we do not review herein,

There are in fact a number of institutions considered as public enterprises by Shepherd, for example, which would escape our definition. Institutions like prisons, public elementary and secondary schools, and certain military production facilities are all government owned, but since they do not sell their output to consumers outside the government we do not regard them as public enterprise. Except for a few remarks that will surface of a comparative institutional nature, we leave consideration of these other institutional forms to one side. (We do, however, give brief attention to quasi-governmental organizations in our final section, which broadly considers recent trends related to public enterprise.)

3.   NORMATIVE DISCUSSION OF PUBLIC ENTERPRISE

Fifty years ago the dominant approach to discussions of public enterprise was very much in the normative realm. At that time mainstream economists were generally predisposed toward government enterprise once one or more so-called market imperfections had been identified (Shleifer 1998, pp. 133-134). Through the 1960s and into the 1970s, economic theorists wrote about the phenomenon of public enterprise in what now can best be described as abstract and overreaching terms. During this period the emphasis was on a "theory" of public enterprise, one distant and detached from fact. For example, R. Turvey writes without apology (1968, p. 7): "what interests us about public enterprise is how it ought to behave. Thus we are not so much concerned with understanding its behavior and making predictions as with criticizing and making recommendations."

Out of this literature come prominent examples of situations in which public enterprises were hypothesized as emerging to address alleged market failures. Shepherd, for instance, lists eight circumstances under which public enterprise can be used to address a problem (1988, pp. 367-368):

- *External effects:* For relieving harms or causing benefits to society which would be overlooked by a private firm. That is, spillover costs and benefits could make private provision unprofitable (e.g., comprehensive school systems, railroad systems).
- *Monopoly:* For either neutralizing current monopolies, fostering competition in oligopolies, or applying vertical countervailing power to a monopoly firm (e.g., utilities).
- *Inadequate private supply:* For correcting the profit motive leading private firms to provide lower than socially optimal levels of output, or the public enterprise acting as a public guarantor for risky new innovations.
- *Inadequate supply to needy users:* For moving the provision of goods and services more into line with ideals of "social fairness" (e.g., universal health care, public housing).
- *Inner nature of the firm:* For modifying the power structure and working conditions of a firm, rendering it more "socially acceptable."
- *Social preference:* For satisfying or mollifying a preference among members of a society, based on culturally determined or exogenous other reasons, for public rather than private provision. This is offered as an explanation of the difference in extent of public enterprise from location to location.
- *Sovereignty:* For protecting a social preference for the exclusion of foreign competitors (e.g., national airlines).
- *Salvage of failing firms:* For bailing out sick industries, deemed important enough to rescue.

Shepherd's list is speculative in nature: there is nothing that lets one know whether these are the actual motivations for public enterprise. Indeed, Shepherd fails to cite specific examples for several of the suggested motivations.

Thomas (Borcherding, 1983) presents a very similar list. Though Borcherding was attempting to articulate a positive theory of public supply, he touches on the body of normative literature. Citing a Kirsch and Yale survey of Canadian Crown government enterprises, Borcherding (1983, p. 149) lists seven proximate normative motives for direct government provision:

- *Cultural and political cohesion:* (e.g., national broadcast companies).
- *Protection of jobs:* (e.g., Sydney steel).
- *Development of key sectors:* especially underdeveloped ones (e.g., SynCrude).

- *Realization of scale economies/control of natural monopolies:* (e.g., Ontario Hydro).
- *Price stabilization:* (e.g., various agricultural boards).
- *Security of supply of crucial activities:* (e.g., PetroCan).
- *Distribution of commodities with social costs:* (e.g., liquor, lottery).

Both of these lists are noteworthy for several reasons. First and foremost, they are good illustrations of the goals and approaches of the normative literature on public enterprise. They share some elements, pointing to important recurring themes. However the thing most important about these normative lists is what they *are not.* They are not, and do not approach, a positive theory of the origins of public enterprise. They offer no explicit testable hypotheses and are thus of limited predictive value.

Also striking about the normative literature is the emphasis on the failings of private firms. Generally, the normative literature has advocated public enterprise in circumstances where the private provision of the good is somehow flawed, whether in terms of efficiency or some abstract ideas of 'fairness.' This begs the question of whether there is a good reason to advocate public provision in its own right, independent of the status of private provision of the good.

One must consider also the generality of the normative arguments, as highlighted by these two lists. While there is an effective argument that in their given circumstances, public enterprise *may* arise, neither list of normative motives excludes other solutions for the situations presented. That is, the normative literature does not present any situations for which public enterprise is the only (or the best) way of dealing with the alleged problem. For example, while public enterprise is certainly one option for coping with the perception of an inadequate supply of a good by private, for-profit firms, it is by no means the only option. Subsidizing private firms, or for that matter consumers themselves (such as has been done in the case of solar panels in the U.S.), could very well achieve the desired result, and possibly for a lower cost, as could a mixed enterprise. In fact the tools of regulation, subsidization and mixed enterprise are often utilized. But, as mentioned earlier in the paper, the literature relevant to each of these areas falls outside our definition of public enterprise and is thus not dealt with here. Instead we want to consider, given the array of possible solutions to perceived problems in resource allocation, why public enterprise? Why would the choice be made for direct government provision?

By contrast, V. V. (Ramanadham, 1991) offers two lists of reasons for public enterprise: one theoretical, in the vein of what we have been calling the normative literature, and one compiled from official government positions. Differing only somewhat from the lists offered by Shepherd and Borcherding, the following theoretical list comes from Ramanadham (1991, pp. 5-20):

- *Growth aspects:* For achieving a desired overall rate of economic growth.
- *Distributional aspects:* For encouraging the theme of "distributional justice," especially by reducing the problems of monopoly, market and wealth concentration, and prevalence of foreign capital, especially in developing economies.
- *Surplus argument:* For generating surpluses to remain in the public sector, for use by the government in expenditures and investments.
- *Comparative advantage:* For utilizing the comparative advantage of organizing a given enterprise in the public sphere, rather than the private sphere.

More remarkable is the second list from Ramanadham (1991, p. 22-37), detailing the reasons or explicit goals for various public enterprises at their inceptions as offered by government spokespersons in various countries:

- *Plan strategy and social gain.*
- *Status of private entrepreneurship.*
- *Aid to private enterprise.*
- *Control over the economy.*
- *Deconcentration.*
- *Anti-monopoly.*
- *Ownership and social restructuring.*
- *Distributional justice.*
- *Savings for investment.*
- *Special considerations.*
- *Agency of development.*

Despite the officially stated rationales appearing in numerous government reports, as cited in Ramanadham, we are skeptical as to whether they are the true underlying motivations to public enterprise—and with good reason. Ramanadham most carefully documents that public enterprises have not achieved their publicly stated goals; there is in fact little connection between the stated rationales for public enterprises and their actual performance. Considering the shortcomings and inefficiencies of the political process from which public enterprises emanate, which we describe later in a section on public choice theory, the question remains: what truly motivates political actors to create a public enterprise? Through Ramanadham we know that market imperfection is the putative reason for creation of public enterprise, but there are other issues that demand consideration.

We shift now from the normative literature to a positive approach to explaining the real world occurrences of public enterprise. While the normative, theoretical approach is perhaps useful to a discussion of a hypothetical public enterprise, we must shift to a positive approach to explain actual public enterprise. We first review the property rights approach to a positive theory of

public enterprise and subsequently propose an alternative approach based on public choice theory.

4.   THE PROPERTY RIGHTS PARADIGM AND EMPIRICAL STUDIES OF PUBLIC ENTERPRISE

There exists an extensive literature in the property rights realm on the nature of public enterprise. The central thesis to this property rights paradigm is that the most important difference between private and public firms is the lack of transferable ownership rights in the latter.

A citizen/taxpayer is a part owner of any enterprise owned by the government to which taxes are paid or within which residence occurs (whether municipal, state or federal). The share of ownership is a function of the number of constituents within that governmental unit. The only way to divest of ownership is to change one's location of home or work, and increasing one's proportionate share of ownership is not possible in a one-person, one-vote state. This makes specialization of ownership in a public enterprise impossible, which in turn limits and inhibits the incentive of any citizen/taxpayer/owner to monitor the behavior of the enterprise.

Owners of private firms, on the other hand, do have an incentive to incur the costs of monitoring the actions of management, according to property rights theory. Since ownership shares of private firms can be concentrated, specialization in ownership occurs, and owners are induced to incur the cost of monitoring their managers. Concentrated ownership makes monitoring of management economical and close monitoring induces managers to maximize the wealth of the private owners. One owner (i.e., citizen) of a public firm has less incentive to incur monitoring costs since any additional wealth spawned by that investment will be dispersed among all owners. Thus, public managers will not act to maximize the wealth of the owners. They will have far more discretion than their private counterparts to shift resource use toward their own personal benefit. They will produce at higher costs with relative ease because no citizen (i.e., owner) will find it worthwhile to try to alter the behavior. Illustration on this point can be found in Table 1 in Vining and Boardman (1992, p. 208), not reproduced here. The survey of the relevant empirical literature presented there reveals that a majority of studies have found that owner controlled firms are more efficient than manager controlled firms (such as public enterprises) when measured in terms of profitability, growth, profit variability, and expenses.

There has been extensive empirical work comparing the relative efficiency of public and private enterprises. The range of industries that have been studied is vast—electric utilities, refuse collection, water, health services, airlines, railroads and other transportation, financial institutions and fire services—and

the results of those studies do not form a tight consensus. Whereas numerous empirical studies have concluded that private companies are more efficient, some others have found no difference in efficiency or even superior efficiency among public enterprises.

Boardman and Vining (1989, p. 4-5) point out several factors that are crucial to explaining the variation in conclusions across empirical studies. First is the fact that different authors utilize different performance and behavior indicators in their studies. Simply, one author's "efficient" may not be the same as another's. The problem is illustrated by considering the yardstick of 'profitability.' In competitive circumstances, profitability can be an appropriate measure of internal efficiency. Some argue, however, that the comparative empirical studies which utilize a profitability comparison ignore the fact that the (general) lower profits of public firms are indicators, not of inefficiency, but of unmeasured sociopolitical outputs. Since the empirical studies do not measure the external benefits of public enterprises, they are unfair comparisons (Boardman and Vining, 1989, p. 8-9).

Results also differ from sector to sector. In industries with very limited competition or strong regulation of private firms, there is more support for the superiority of public firms; in industries where governments subcontract to the private sector, private enterprises seem to be the more efficient service provider.

Vining and Boardman (1992, p. 214-215) summarize the results of the empirical studies of relative efficiency in their Table 2, which we reproduce below as Table 1. The table is an update of one appearing in Boardman and Vining (1989, p. 6). Ignoring the differences in measurement outlined above, it indicates that the overall empirical results seem to favor private enterprise.

Boardman and Vining (1989, p. 1-8), however, argue that the extant empirical work does not provide strong support for the property rights hypothesis that private enterprise should be more efficient than public enterprise. One reason is the early property rights literature failed to recognize that the act of disciplining corporate management is inhibited by a free rider problem. Takeovers of efficient firms raise share prices, but to the extent that existing shareholders hold onto their shares to benefit from the takeover they inhibit the takeover from being accomplished (Grossman and Hart, 1980). Another problem addressed by Boardman and Vining is that much empirical work is limited in value by the fact that the evidence comes from firms that enjoy a natural monopoly, or operate in a duopoly, or are prevented from operating at competitive prices by government regulation. Taking issue with the lack of research on the effects of ownership in a competitive environment, Boardman and Vining control for such factors in their own empirical study. Testing the effect of ownership on efficiency, while controlling for competitive circumstances, their study finds that private enterprises operate significantly more

TABLE 1.
Empirical Results on Relative Efficiency of Public and Private Corporations

Sector	Public corporation more efficient	No difference or ambiguous results	Private corporations more Efficient
Electric utilities	(Meyer, 1975) (Neuberg, 1977) (Pescatrice and Trapani, 1980)	(Mann, 1970) (Dilorenzo and Robinson, 1982) (Färe et al., 1985) (Atkinson and Halvorsen, 1986)	(Shepherd, 1966) (Moore, 1970) (Peltzman, 1971) (Tilton, 1973) (De Alessi, 1974) (De Alessi, 1977)
Refuse	(Pier et al., 1974)	(Hirsch, 1965) (Kemper and Quigley, 1976) (Collins and Downes, 1977)	(Kitchen, 1976) (Savas, 1977) (Pommerehne and Frey, 1977) (Spann, 1977) (Stevens, 1978) (Edwards and Stevens, 1978) (Boorsma, 1982) (McDavid, 1985) (Lawarrée, 1986)
Water	(Mann and Mikesell, 1971) (Bruggink, 1982)	(Feigenbaum and Teeples, 1983) (Lindsay, 1984)	(Hausman, 1976) (Morgan, 1977) Crain and Zardkoohi (1980, 1980) (Boland, 1983)
Health related services, nursing homes, and daycare	(Pattison and Katz, 1983)	(Becker and Sloan, 1985) (Renn et al., 1985) (Tuckman and Chang, 1988)	(Clarkson, 1972) (Hrebiniak and Alutto, 1973) (Rushing, 1974) (Lindsay, 1976) (Hsaio, 1978) Frech (1976, 1980, 1985) (Bays, 1979) (Bishop, 1980) (Frech and Ginsburg, 1981) (Finsinger, 1982) (Wilson and Jadlow, 1982) (Sloan and Vraciu, 1983) (Lee et al., 1983) (Schlesinger and Dorwart, 1984) (Schulz et al., 1984) (Herzlinger and Krasker, 1987) (Mauser), 1988) (Tuckman and Chang, 1988)
Airlines		(Forsyth and Hocking, 1980) (Morrison, 1981) (Jordan, 1982)	(Davies, 1971), (Davies, 1977) (Mackay, 1979), (Pryke, 1982) (Findley and Forsyth, 1984) (Kirby, 1986) (Forsyth, Hill, and Trengove, 1986) (Gillen, Oum, and Tretheway, 1989)
Railroads		(Caves and Christensen, 1980) (Caves et al., 1982) (Freeman et al., 1985)	
Financial institutions		(Lewin, 1982)	(Davies, 1981) (Davies and Brucato, 1987)
Fire services			(Ahlbrandt, 1973)
Non-rail transit			(Pashigian, 1976), (Bays, 1979) (Pryke, 1982), (Pucher, 1982) (Pucher et al., 1983)

TABLE 1.
Continued

Sector	Public corporation more efficient	No difference or ambiguous results	Private corporations more Efficient
			(Palmer et al., 1983) (McGuire and Van Cott, 1984) (Wallis, 1985) (Perry and Babitsky, 1986) (Hensher, 1987)
Misc. industrial			(Funkhouser and MacAvoy, 1979) (Kim, 1981) (Rowley and Yarrow, 1981) (Ayub and Hashimoto, 1985) (Boardman and Vining, 1989) (Picot and Kaulmann, 1989)

efficiently than public enterprises, supporting the (amended) property rights conjecture (1989, p. 26).

Although they critique the method of previous property rights economists, Boardman and Vining reach a similar conclusion. Moreover, the ownership/efficiency hypothesis of the property rights school, amended but ultimately upheld with Boardman and Vining's strong empirical support, is still of limited usefulness. Specifically, even though it adequately addresses performance according to narrow economic definitions of efficiency, the property rights literature is silent with regards to several important issues concerning the emergence of, and motivations for public enterprise. In short, the property rights approach, while useful to certain issues, fails to achieve the degree of scope that the public choice analysis that follows does.

Some studies overlap the property rights and public choice approaches, which makes classification difficult. Using the same dichotomous categories, Borcherding (1983, p. 103, 125) recognized that what is referred to as a property rights analysis as opposed to a public choice analysis is at times somewhat artificial and arbitrary. One guiding principle we use in placing a study in the property rights camp is to do so if the study focussed narrowly on the issue of comparative efficiency of public versus private firms. If the range of issues addressed was wider than that, the study is considered under public choice.

5. PUBLIC CHOICE AND PUBLIC ENTERPRISE

Several reasons motivate the approach here. At a minimum, it seems to us that the spirit of public choice analysis leads to questions we consider vital to an understanding of the subject at hand: *What factors actually induce public enterprises; how are public enterprises in fact run; for whose benefit; and at whose expense?* Public choice offers a recognized, well-developed and unified theory which guides the analysis toward answers to our specific set of key

questions. As well, public choice analysis is the natural handmaiden of the constitutional perspective characterizing the larger work of which this chapter is but one part. The approach here is thus justifiable both intrinsically and instrumentally.

Now for the bad news: even adhering to our limited definition of what constitutes public enterprise, we do not expect to use public choice analysis here to yield answers to our set of key questions for the multitude of examples of public enterprise in its varied forms throughout the world. In fact, at this juncture we further limit our inquiry to public enterprise in the context of representative democracy for the simple reason that public choice theory has been most finely tuned to exploring the consequences of representative democracy, such as in the U.S. This point of departure is no small matter for a chapter that thus far has considered public enterprise globally. Our rationale is simple: the instrument of analysis must be suitably calibrated to the task at hand, and we utilize the instrument where it has the best chance of successful application; specifically, we use it in the context in which it has received the greatest refinement in the extant literature. In so doing, we do not mean to suggest that the lens of public choice is ill-suited to application in contexts other than representative democracy, only that it would need re-calibration to the institutional framework at play in other contexts.

We find ourselves on a pathway to a theory, with no pretensions that the framework is sufficiently well developed to answer our set of key questions for every real world instance of public enterprise. Instead, we regard the framework as useful for highlighting ubiquitous factors that need to be accounted for in seeking answers to our key questions, regardless the particular example of public enterprise. In this sense, we do indeed think of public choice analysis as a lens of inquiry [cf. (Crew and Rowley, 1988) and (Rowley, 1995) on the subject of a public choice theory of regulation). In addressing specific instances of public enterprise, however, the general theory must be used in conjunction with suitable knowledge of whatever idiosyncratic forces are also at play (for instance the contribution of technological change toward explaining the shifting extent of government enterprise in the U.S. electric utility industry (Schap 1986)).

Public choice theory recognizes that ordinary citizens cannot readily transform their political preferences into actual policy outcomes in a representative democracy. Politicians serve as brokers of wealth transfers. Those voters who are able to coalesce into effective interest groups seek such transfers, while other, unorganized citizens incur the associated costs. Benefits are concentrated to the special interests, with resultant costs widely dispersed among the unorganized. Politicians attempt to conceal and (given election cycles) postpone the true costs of initiatives. Each voter is rationally ignorant on most

issues because it is simply not worthwhile to become informed. Within an in-stitutionalized set of rules, each politician balances competing interest groups toward an ultimate goal of maximizing the politician's own wellbeing, broadly considered. Politics in certain circumstances can become a negative sum game because of rent-seeking leakage—competition for wealth transfers results in social waste of resources. To identify competing interests and political con-straints, then, is to begin to understand the motivations for public enterprises, the ways in which they will operate and the resultant gainers and losers. More elaborate (and more carefully referenced) summary presentations of the pub-lic choice approach can be found in the literature [e.g., Gwartney and Stroup (1997) and Gwartney and Wagner (1988)], but the preceding few sentences cover the essential elements of the approach and suffice for our purposes here.

The choice to enter into public enterprise is an institutional choice that im-plicitly rejects other market structures. These include provision (either by mar-ket forces or by contracting out by the state) by private, for-profit firms; not-for-profit firms; worker-managed firms; regulated, for-profit firms; and mixed enterprise. If the output could conceivably be allocated by a means other than by sale, then agency and bureaucracy enter the picture as well. The decision for public enterprise, then, is not an absolute one made in a vacuum, but is instead shaped and tempered by the existence of alternative institutional possibilities. (Borcherding, 1983) shares this view.

A public choice framework of analysis is presented in tabular form below (Table 2). Sub-categories are not mutually exclusive. Our brief discussion of the framework is suggestive rather than exhaustive. For each of the three main

TABLE 2.
Public Choice Framework

Actors	Phenomena	Constraints
• POLITICIANS	• WEALTH TRANSFERS/ CROSS SUBSIDIZATION	• FINANCIAL
• CITIZENS		• CIVIL SERVICE
• VOTERS	• COALESCENCE OF INTERESTS	• FEDERALISM
• TAXPAYERS		• VOTING (CAMPAIGN CONTRIBUTION LIMITS, COMMITTEE SYSTEM, AGENDA CONTROL)
• CONSUMERS OF OUTPUT	• TRANSACTIONS COSTS	
• ENTERPRISE MANAGERS	• RATIONAL IGNORANCE	
• ENTERPRISE WORKERS	• CONCENTRATED BENEFITS/ DISPERSED COSTS	
• SELLERS OF INPUTS		
• COMPETITORS AND THEIR EMPLOYEES	• SHORTSIGHTEDNESS	
• POTENTIAL COMPETITORS	• COST CONCEALMENT	
	• REDISTRIBUTION CONCEALMENT	
	• IMPERFECT OVERSIGHT	

categories, we present glimpses of the public choice implications that can be drawn for the more prominent sub-categories, often with citation to published works that provide more detailed accounts.

Concerning the relevant actors, the behavior of politicians, taxpayers, and consumers has been briefly described already. Here we simply add the crucial point that consumers must somehow coalesce if they are to influence policy. Interestingly, the public enterprise may have to come into being before consumer interests can identify each other at sufficiently low cost to make formation of a consumer interest group feasible. In this way, a public enterprise can overtime create its own constituency. This suggests the importance of an intertemporal analysis of public enterprise. The story of public enterprise workers also features dynamic aspects. Once organized, worker interests will attempt to influence to their benefit the input mix within the public enterprise, which in turn will affect the internal efficiency of the enterprise over time. Depending on the degree of political oversight, public enterprise managers also have complex incentives to tinker with the capital-labor mix within the enterprise (Zardkoohi and Giroux, 1990).

Certain phenomena highlighted in public choice analysis warrant elaboration. The notion of transaction costs, for example, is multifaceted in this context. At one level of analysis transaction costs can be seen as affecting the feasibility of interest-group formation; at another level, they are observed limiting or inhibiting government use of "contracting out" with private firms as an alternative to government enterprise. Similarly, the phenomenon of a general shortsightedness of political decision-makers, driven by impending reelection concerns, manifests itself in multiple ways in the operations of public enterprises. A dynamic analysis of public enterprise, founded on the phenomenon of political myopia, would note a tendency toward cost postponement by virtue of deferred maintenance and repair of physical plant. Wealth redistribution favors current citizens/consumers through artificially low current prices; a legacy of higher future prices is left to future generations of consumers/voters whose unorganized interests are not represented at present in the political process (Schap, 1988). Political oversight of public enterprise is not perfect, however, and must be done by proxy in the absence of a pure profit motive. Consequently, managers of public enterprises can be expected to concentrate importance on output characteristics that are highly visible while de-emphasizing those attributes which, though important, are less easily monitored. An example, in the context of a Veterans Administration hospital, would be the difference between easily counting the number of patient-days provided versus the difficulty of evaluating the quality of physician bedside manner. Relatively more of the former characteristic will be provided by a public enterprise than would be provided by a for-profit hospital (Lindsay, 1976).

Certain legal constraints, some of which are legislated and others constitutional, need to be incorporated in any public choice analysis of public enterprise. Chief among these, for the purposes of this volume, is the set of financial constraints. One prominent financial issue is whether the public enterprise operates entirely off budget. If so, despite whatever socioeconomic benefits it may provide, the enterprise will be constrained to cover its costs. When a public enterprise is subsidized, those subsidies compete squarely in the budgetary process with other government projects and activities for scarce dollars. When a budget constraint exists, the competition for scarce dollars is heightened. A profitable public enterprise can be required to contribute to the general fund, and the decisions concerning how much and how to use the revenue become interesting objects of public choice analysis (Deno and Mehay, 1988). Political decisions concerning which activities get treated as off-budget (Marlow and Joulfaian, 1989) influence the creation and ultimate performance of public enterprises.

6.   RECENT DEVELOPMENTS RELATED TO PUBLIC ENTERPRISE

Apart from the collapse of communism, the widespread privatization of public enterprises during the 1980s and 1990s is the most obvious and important recent development in the U.S., U.K., and elsewhere. We ignore privatization here despite its importance because it is the focus of another entire chapter in this volume. Here we concentrate instead on two other interesting recent developments related to public enterprise, one largely European and the other American.

6.1.   Quangos

Although the word *quango* is not well known among economists and policy makers on the West Side of the Atlantic, it has entered and become entrenched in the public discourse of several European countries—especially in the United Kingdom. The reason for this has to do with the preponderance of governmental power held in the hands of the executive branch in many European countries.

While the term quango itself stands for quasi-governmental organizations, its meaning has been defined as any private or public body "that spends public money to fulfil a public task but with some degree of independence from elected representatives" (Flinders, 1999a, p. 4). Since each verb and adjective above is subject to levels or degrees of intensity, the word conveys a variety of meanings that can only be deciphered within specific institutional environments. At the national level in the United Kingdom, for example, these organizations include advisory bodies to the national government, tribunals

performing quasi-judicial functions, the Boards of Visitors that oversee the penal system, and executive organizations, which manage some government programs. However, at the local and regional levels, quangos include housing associations, health care trusts, police authorities, grant-maintained schools, and the local government councils (Flinders, 1999a, p. 5).

Quangos, then, are not strictly speaking agencies—because civil service rules do not apply to them and the managers are supposed to be independent of the bureaucracy and elected politicians; yet, the power of the purse may be used by the national government to influence their behavior. Besides, the managers of these organizations are appointed by the body politic, and the organizations have been used to reward members of the party in power. In addition, quangos have been created to circumvent local authorities that are not responding to the wishes of the national government (Flinders, 1999a, p. 9).

Neither are quangos government enterprises, at least not according to our strict definition given earlier, for they do not generally sell goods or services to the public. But since they function in many cases as supervisors and regulators of privatized companies (such as gas, water, the railways, the lottery, etc.) one cannot ignore the fact that they do play a role (even if an indirect one) in the production of goods and services.

The strengths and weaknesses of quangos are almost one and the same. Their independence, if successfully exercised by aggressive management, may in fact serve a public purpose that neither the national nor local governments would be willing to pursue. But that same independence allows them to maintain a level of secrecy and, in some cases, unaccountability that makes citizens think of them as illegitimate and corrupt (Flinders, 1999b).

In the same way that there is no widely accepted positive theory of government enterprises, there is great puzzlement as to why quangos have come into prominence over the last two decades. Some have argued that by passing on governmental functions to supposedly independent entities it is possible for politicians to both claim credit for reducing the size and power of the bureaucracy and simultaneously entrench power in new experts who over time become less accountable to anyone but themselves (Landers, 1999). However, this assertion does not deny that occasionally, or even often, quangos look after what many perceive to be the public interest, and in so doing may perform a useful function.

In the United States, the Federal Reserve System may be thought of as one of those quangos over which the legislative and executive branches of government exert some degree of supervision. But no one would deny that the recent effectiveness of the System depends on the ability of the Chairman of the Board of Governors to exercise autonomy in situations when political considerations push the System towards the satisfaction of narrow interests.

6.2.   Public Enterprise via Asset Seizure

Legal changes enacted in the 1980s launched many episodes of government enterprise in the United States of a kind that could not reasonably be anticipated based on a reading of the extant economics literature concerning public enterprise. Cases of government enterprise arose arguably by accident, the unintended consequence of modern asset forfeiture law. Law enforcement authorities at the federal level and in many state and local jurisdictions as well during the 1980s were granted expanded opportunities to seize assets used in the commission, or conspiracy to commit, certain criminal acts (principally, but not exclusively, racketeering, terrorism and narcotics trafficking). In such cases, asset seizure would proceed without first proving at trial guilt "beyond a reasonable doubt," and even a subsequent verdict of "not guilty" would not necessitate the return of previously seized property. As a matter of fact, and of the law for a period of time, the mere allegation of criminal wrongdoing, supported only by the minimal legal standard of "probable cause" on the part of enforcement authorities, would subject one's property to asset seizure by those same authorities. Worse yet for an innocent property owner was the fact that the return of seized property was made contingent on a demonstration, by the hefty legal norm of "preponderance of evidence," that the property was not used in the commission of a crime. The guilt or innocence of the property owner with respect to the alleged crime was irrelevant to the issue of the return of previously seized property: the property itself, not its owner, was held guilty! Hyde (1995, pp. 17-27) describes the evolution of this remarkable legal doctrine.

Authorities seized entire business enterprises with some regularity when one or more of the assets of the business had allegedly been used in a crime. Once seized, all or some of the business assets would become available to fund future activities of the authorities. Not surprisingly, the opportunity to seize assets, which could then be used to fund the operations of the very same authorities doing the seizing, led to substantial abuse (Hyde, 1995, pp. 29-53). Our purpose here is not so much to condemn abuse of power asrit is merely to call attention to this novel, indirect source of government enterprise.

Once assets were seized, the authorities would have an obvious interest in preserving their value until such time as they could be sold or auctioned off, either piecemeal or in toto. Since the value of a business as an ongoing concern can far exceed the value of its liquidated assets, value preservation became a motivation and rationalization for law enforcement authorities undertaking the day-to-day operations of very ordinary businesses. No mere flash experiences, some seized businesses were operated by law enforcement authorities for periods of time measured in years. One particularly notorious example (reported on the television program "American Justice") involved a seized casino

that was totally remodeled while under government operation for a number of years; government officials even went so far as to hire a new manager away from one of the competing casinos! Numerous other opportunities for side-door entry into government enterprise have been reported (Hyde, 1995, pp. 29-53) and include real estate (apartment houses, fraternity houses, hotel and golf courses), transportation (air charter service, shipping), banking, retail car sales (dealership), mining (gold), entertainment (movie theater), and retail sales (electronics and convenience stores). And lest the reader think that this particular list is exhaustive with respect to possible business type, once seizure for tax fraud is contemplated there is no line of business that could be thought of as categorically insulated from government enterprise instigated by asset seizure. Public enterprises of various kinds could occur seemingly wherever the wind blew.

After more than a decade of what came to be perceived as substantial abuse, the practice of asset seizure was addressed by legislation enacted in the U.S. Congress in April 2000 by unanimous voice vote. The legislation was designed to sharply curtail the practice of asset seizure by federal authorities. A full appraisal of the ultimate impact of the legislation must await its interpretation by the courts.

7.   SUMMARY

The literature on public enterprise is vast. We have attempted to present some of its major themes in a manageable package. We have highlighted the issues involved in defining public enterprise, identified ways in which public enterprises have been characterized in the literature, summarized empirical research on the subject, and noted important recent developments related to public enterprise. Our section on public choice and public enterprise departed from mere retrospective review by suggesting prospectively that public choice theory can furnish the basis for a general framework of positive analysis of public enterprise. Indeed, we have offered an itemized list of what such a framework of analysis should include and account for. Although many published articles describe one or more aspects of public enterprise the literature still awaits the appearance of a comprehensive and generally accepted theory of public enterprise.

REFERENCES

Aharoni, Y. (1986). *The Evolution and Management of State Owned Enterprises.* Cambridge, MA: Ballinger Publishing Co.

Ahlbrandt, Jr., R. (1973). "Efficiency in the Provision of Fire Services." *Public Choice* 16 (Fall): 1-15.

Alchain, A. A. (1965). "Some Economics of Property Rights." *Il Politico* 30: 816-829. Reprinted in *Economic Forces at Work,* Alchian, A. A., pp. 127-149. Indianapolis: LibertyPress, 1977.

Atkinson, S. E. and Halvorsen, R. (1986). "The Relative Efficiency of Public and Private Firms in a Regulated Environment: The Case of U.S. Electric Utilities." *Journal of Public Economics* 29 (April): 281-294.

Ayub, M. A. and Hashimoto, H. (1985). *The Economics of tin Mining in Bolivia.* Washington, DC: The World Bank.

Bays, C. W. (1979). "Cost Comparisons of Forprofit and Nonprofit Hospitals." *Social Science and Medicine* 13C (December): 219-225.

Becker, E. R. and Sloan, F. A. (1985). "Hospital Ownership and Performance." *Economic Inquiry* 23 (January): 21-36.

Bishop, C. E. (1980). "Nursing Home Cost Studies and Reimbursement Issues." *Health Care Financing Review* 1 (Spring): 47-64.

Boardman, A. E. and Vining, A. R. (1989). "Ownership and Performance in Competitive Environments: A Comparison of the Performance of Private, Mixed and State-owned Enterprises." *Journal of Law and Economics* 32 (April): 1-33.

Boardman, A. E. and Vining, A. R. (1990). "The Behavior of Mixed Enterprises." *Journal of Research in Law and Economics* 14: 223-250.

Boland, J. J. (1983). *Water and Wastewater Pricing and Financial Practices in the United States.* Washington, DC: Metametrics.

Boorsma, P.B. (1982). "Public Sector Productivity and Relative Efficiency in the Netherlands." Unpublished paper, Department of Public Administration, Twente University.

Borcherding, T. E. (1983). "Toward a Positive Theory of Public Sector Supply Arrangements." In *Crown corporations in Canada,* Prichard, J. R. S., ed., pp. 99-184. Toronto, Ontario: Buttersworth.

Bös, D. (1989). *Public Enterprise economics: Theory and Application,* 2nd ed. New York: North Holland.

Bruggink, T. H. (1982). "Public Versus Regulated Private Enterprise in the Municipal Water Industry: A Comparison of Operating Costs." *Quarterly Review of Economics and Business* 22 (Spring): 111-125.

Caves, D. W. and Christensen, L. R. (1980). "The Relative Efficiency of Public and Private Firms in a Competitive Environment: The Case of Canadian Railroads." *Journal of Political Economy* 88 (October): 958-976.

Caves, D. W., Christensen, L. R., Swanson, J. A., and Tretheway, M. W. (1982). "Economic Performance of U.S. and Canadian Railroads: The Significance of Ownership and the Regulatory Environment." In *Managing public enterprises,* Stanbury, W. T. and Thompson, F., eds., pp. 123-151. New York: Praeger.

Clarkson, K. W. (1972). "Some Implications of Property Rights in Hospital Management." *Journal of Law and Economics* 15 (October): 363-384.

Collins, J. N. and Downes, B. T. (1977). 'The Effect of Size on the Provisions of Public Services: The Case of Solid Waste Collection in Smaller Cities." *Urban Affairs Quarterly* 12 (March): 333-345.

Crain, W. M. and Zardkoohi, A. (1980). "Public Sector Expansion: Stagnant Technology or Attenuated Property Rights?" *Southern Economic Journal* 46 (April): 1069-1082.

Crain, W. M. and Zardkoohi, A. (1978). "A Test of the Property Rights Theory of the Firm: Water Utilities in the United States." *Journal of Law and Economics* 21 (October): 395-408.

Crew, M. A. and Rowley, C. K. (1988). "Toward a Public Choice Theory of Monopoly Regulation." *Public Choice* 57: 49-67.

Davies, D. G. (1971). "The Efficiency of Public Versus Private Firms: The Case of Australia's Two Airlines." *Journal of Law and Economics* 14 (April): 149-165.

Davies, D. G. (1977). "Property Rights and Economic Efficiency—The Australian Airlines Revisited." *Journal of Law and Economics* 20 (April): 223-226.

Davies, D. G. (1981). "Property Rights and Economic Behavior in Private and Government Enterprises: The Case of Australia's Banking System." In *Research in Law and Economics,* Zerbe, R. O. ed., Vol. 3, pp. 111-142. Greenwich, CT: JAI Press.

Davies, D. G. and Brucato, Jr., P. F. (1987). "Property Rights and Transaction Costs: Theory and Evidence on Privately-owned and Government-owned Enterprises." *Journal of Institutional and Theoretical Economics* 143 (March): 7-22.

De Alessi, L. (1974). "An Economic Analysis of Government Ownership and Regulation: Theory and the Evidence from the Electric Power Industry." *Public Choice* 19 (Fall): 1-42.

De Alessi, L. (1977). "Ownership and Peak-load Pricing in the Electric Power Industry." *Quarterly Review of Economics and Business* 17 (Winter): 7-26.

De Alessi, L. (1980). "The Economics of Property Rights: A Review of the Evidence." *Research in Law and Economics* 2: 1-47.

De Alessi, L. (1982). "On the Nature and Consequences of Private and Public Enterprises." *Minnesota Law Review* 67 (October): 191-209.

Deno, K. T. and Mehay, S. L. (1988). "Municipal Utilities and Local Public Finance: A Simultaneous Model." *Public Choice* 57: 201-212.

Dilorenzo, T. J. and Robinson, R. (1982). "Managerial Objectives Subject to Political Market Constraints: Electric Utilities in the U.S." *Quarterly Review of Economics and Business* 22 (Summer): 113-125.

Edwards, F. R. and Stevens, B. J. (1978). "The Provision of Municipal Sanitation by Private Firms: An Empirical Analysis of the Efficiency of Alternative Market Structures and Regulatory Arrangements." *Journal of Industrial Economics* 27 (December): 133-147.

Färe, R., Grosskopf, S. and Logan, J. (1985). "The Relative Performance of Publicly-owned and Privately-owned Electric Utilities." *Journal of Public Economics* 26 (February): 89-106.

Feigenbaum, S. and Teeples, R. (1983). "Public versus Private Water Delivery: A Hedonic Cost Approach." *Review of Economics and Statistics* 65 (November): 672-678.

Findley, C. C. and Forsyth, P. J. (1984). "Competitiveness in Internationally Traded Services: The Case of Air Transport." ASEAN-Australia Joint Research Project Working paper.

Finsinger, J. (1982). "The Performance of Public Enterprises in Insurance Markets." Paper presented at a *Conference on the Performance of Public Enterprise,* Liege, Sart Tilman, Belgium.

Flinders, M. V. (1999a). "Setting the Scene: Quangos in Context." In *Quangos, Accountability and Reform: The Politics of Quasi-government,* Flinders, M. V. and Smith, M. J., eds., pp. 3-16. Basingstoke: Macmilian.

Flinders, M. V. (1999b). "Quangos: Why Do Governments Love Them?" In *Quangos, Accountability and Reform: The Politics of Quasi-government,* Flinders, M. V. and Smith, M. J., eds., pp. 26-39. Basingstoke: Macmilian.

Forsyth, P. J. and Hocking, R. D. (1980). "Property Rights and Efficiency in a Regulated Environment: The Case of Australian Airlines." *The Economic Record* 56 (June): 182-185.

Forsyth, P. J., Hill, R. D., and Trengove, C. D. (1986). "Measuring Airline Efficiency." *Fiscal Studies* 7 (February): 61-81.

Frech, III, H. E. (1976). "The Property Rights Theory of the Firm: Empirical Results from a Natural Experiment." *Journal of Political Economy* 84 (February): 143-152.

Frech, III, H. E. (1980). "Property Rights, the Theory of the Firm, and Competitive Markets for Top Decisionmakers." In *Research in Law and Economics,* Zerbe, R., ed., Vol. 2, pp. 49-63. Greenwich, CT: JAI Press.

Frech, III, H. E. (1985). "The Property Rights Theory of the Firm: Some Evidence from the U.S. Nursing Home Industry." *Journal of Institutional and Theoretical Economics* 141 (March): 146-166.

Frech, III, H. E. and Ginsburg, P. B. (1981). "The Cost of Nursing Home Care in the United States: Government Financing, Ownership and Efficiency." In *Health, Economics and Health Economics,* van der Gaag, J. and Perlman, M, eds., pp. 67-81. New York: North Holland Publishing Company.

Freeman, K. D., Oum, T. H., Tretheway, M. W., and Waters, II, W. G. (1985). "The Total Factor Productivity of the Canadian Class 1 Railways: 1956–1981." *The Logistics and Transportation Review* 21 (September): 249-276.

Funkhouser, R. and MacAvoy, P. W. (1979). "A Sample of Observations on Comparative Prices in Public and Private Enterprises." *Journal of Public Economics* 11 (June): 353-368.

Gillen, D. W., Oum, T., and Tretheway, M. W. (1989). "Privatization of Air Canada: Why it Is Necessary in a Deregulated Environment?" *Canadian Public Policy* 15 (September): 285-299.

Grossman, S. J. and Hart, O. D. (1980), "Takeover Bids, the Free Rider Problem and the Theory of the Corporation." *Bell Journal of Economics* 11 (Spring): 42-64.

Gwartney, J. and Wagner, R. E. (1988). "The Public Choice Revolution." *The Intercollegiate Review* 23 (Spring): 17-26.

Gwartney, J. D. and Stroup, R. L. (1997). "Public Choice: Understanding Government and Government Failure." In *Microeconomics: Private and Public Choice,* 8th ed.. Chapter 19, pp. 491-515. Dryden-Harcourt Brace Publishers.

Hausman, J. M. (1976). "Urban Water Services Pricing: Public vs. Private Firms." Ph.D. Thesis, Department of Economics, George Washington University.

Heath, J. (ed.) (1990). *Public Enterprise at the Crossroads: Essays in Honour of V.V. Ramanadham.* New York: Routledge.

Hensher, D. (1987). "Productive Efficiency, Ownership and Contracting out of Transport Services." In *Privatization: An Australian Perspective,* Abelson, P., ed. Sydney, NSW: Australian Professional Publications.

Herzlinger, R. E. and Krasker, W. S. (1987). "Who Profits from Nonprofits?" *Harvard Business Review* 65 (January-February): 93-106.

Hirsch, W. Z. (1965). "Cost Functions of an Urban Government Service: Refuse Collection." *Review of Economics and Statistics* 47 (February): 87-92.

Hrebiniak, L. G. and Alutto, J. A. (1973). "A Comparative Organizational Study of Performance and Size Correlates in Inpatient Psychiatric Departments." *Administrative Science Quarterly* 18 (September): 365-382.

Hsaio, W. (1978). "Public versus Private Administration of Health Insurance: A Study in Relative Economic Efficiency." *Inquiry* 14 (December): 379-387.

Hyde, H. (1995). *Forfeiting our Property Rights: Is Your Property Safe from Seizure?* Washington, D.C.: Cato Institute,

Jordan, W. A. (1982). "Performance of North American and Australian Airlines: Regulation and Public Enterprise." In *Managing Public Enterprises,* Stanbury, W. T. and Thompson, F., eds., pp. 161-199. New York: Praeger.

Kemper, P. and Quigley, J. M. (1976). *The Economics of Refuse Collection.* Cambridge, MA: Ballinger.

Kim, K. S. (1981). "Enterprise Performance in the Public and Private Sectors: Tanzanian Experience, 1970–1975." *The Journal of Developing Areas* 15 (April): 471-484.

Kirby, M. G. (1986). "Airline Economies of 'Scale' and Australian Domestic Air Transport Policy." *Journal of Transport Economics and Policy* 20 (September): 339-352.

Kitchen, H. M. (1976). "A Statistical Estimation of an Operating Cost Function for Municipal Refuse Collection." *Public Finance Quarterly* 4 (January): 56-76.

Landers, B. (1999). "Of Ministers, Mandarins and Managers." In *Quangos, Accountability and Reform: The Politics of Quasi-government,* Flinders, M. V. and Smith, M. J., eds., pp. 120-131. Basingstoke: Macmillan.

Lawarrée, J. (1986). "Une Comparison Empirique des Performance des Secteurs Prive et Public: Le Cas des Collectes d'immondices en Belgique." *Cahiers Economiques de Bruxelles* 109: 3-31.

Lee, A. J., Birnbaum, H., and Bishop, C. E. (1983). "How Nursing Homes Behave: A Multi-equation Model of Nursing Home Behavior." *Social Science and Medicine* 17 (13): 1897-1906.

Lewin, A. Y. (1982). "Public Enterprise, Purposes and Performance: A Survey of Western European Experience." In *Managing Public Enterprises,* Stanbury, W. T. and Thompson, F., eds., pp. 51-78. New York: Praeger.

Lindsay, B. E. (1984). "New Hampshire Water Systems: Environmental and Ownership Considerations." *Water Resources Bulletin* 20: 901-904.

Lindsay, C. M. (1976). "A Theory of Government Enterprise." *Journal of Political Economy* 84 (October): 1061-1077.

MacAvoy, P. W., Stanbury, W. T., Yarrow, G., and Zeckhauser, R. J. (eds.) (1989). *Privatization and State Owned Enterprises: Lessons from the United States, Great Britain and Canada.* Boston: Kluwer Academic Publishers.

Mackay, K. R. (1979). "A Comparison of the Relative Efficiency of Australian Domestic Airlines and Foreign Airlines, Appendix A6.1 of Department of Transport." In *Domestic Air Transport Policy Review,* Vol. 2, Appendices. Canberra: AGPS.

Mann, P. C. (1970). "Publicly Owned Electric Utility Profits and Resource Allocation." *Land Economics* 46 (November): 478-484.

Mann, P. C. and Mikesell, J. L. (1971). "Tax Payments and Electric Utility Prices." *Southern Economic Journal* 38 (July): 69-78.

Marchand, M., Pestieau, P., and Tulkens, H. (1984). "The Performance of Public Enterprise: Normative, Positive and Empirical Issues." In *The Performance of Public Enterprise: Concepts and Measurements,* Marchand, M., Pestieau, P., and Tulkens, H., eds., pp. 3-42. New York: Elsevier Science Publishers (North Holland).

Marlow, M. L. and Joulfaian, D. (1989). "The Determinants of Off-budget Activity of State and Local Governments." *Public Choice* 63: 113-123.

Mauser, E, (April 1988). "The Supply of Center-based Day Care in Massachusetts: Nonprofits versus For-profits." Unpublished paper. Department of Economics, Wellesley College.

McDavid, J. C. (1985). "The Canadian Experience with Privatizing Residential Solid Waste Collection Services." *Public Administration Review* (September/October): 603-604.

McGuire, R. A. and Van Cott, T. N. (1984). "Public versus Private Economic Activity: A New Look at School Bus Transportation." *Public Choice* 43 (1): 25-43.

Meyer, R. A. (1975). "Publicly Owned Versus Privately Owned Utilities: A Policy Choice." *Review of Economics and Statistics* 57 (November): 391-399.

Moore, T. G. (1970). "The Effectiveness of Regulation of Electric Utility Prices." *Southern Economic Journal* 36 (April): 365-375.

Morgan, W. D. (1977). "Investor Owned vs. Publicly Owned Water Agencies: An Evaluation of the Property Rights Theory of the Firm." *Water Resources Bulletin* 13 (August): 775-781.

Morrison, S. (1981). "Properly Rights and Economic Efficiency: A Further Examination of the Australian Airlines." Unpublished paper, Faculty of Commerce and Business Administration, University of British Columbia.

Neuberg, L. G. (1977). "Two Issues in the Municipal Ownership of Electric Power Distribution Systems." *Bell Journal of Economics* 8 (Spring): 303-323.

Palmer, J., Quinn, J. and Resendes, R. (1983). "A Case Study of Public Enterprise: Gray Coach Lines Ltd." In *Crown Corporations in Canada: The Calculus of Instrument Choice,* Prichard, J. R. S., ed., pp. 369-446. Toronto: Butterworths.

Pashigian, B. P. (1976). "Consequences and Causes of Public Ownership of Urban Transit Facilities." *Journal of Political Economy* 84 (December): 1239-1259.

Pattison, R. V. and Katz, H. M. (1983). "Investor-owned and Not-for-profit Hospitals." *The New England Journal of Medicine* 309 (11 August): 347-353.

Peltzman, S. (1971). "Pricing in Public and Private Enterprises: Electric Utilities in the United States." *Journal of Law and Economics* 14 (April): 109-147.

Perry, J. L. and Babitsky, T. T. (1986). "Comparative Performance in Urban Bus Transit: Assessing Privatization Strategies." *Public Administration Review* (January /February): 63-64.

Pescatrice, D. R. and Trapani, III, J. M. (1980). "The Performance and Objectives of Public and Private Utilities Operating in the United States." *Journal of Public Economics* 13 (April): 259-276.

Picot, A. and Kaulmann, T. (1989). "Comparative Performance of Government-owned and Privately-owned Industrial Corporations—Empirical Results from Six Countries." *Journal of Institutional and Theoretical Economics* 145 (June): 298-316.

Pier, W. J., Vernon, R. B., and Wicks, J. H. (1974). "An Empirical Comparison of Government and Private Production Efficiency." *National Tax Journal* 27 (December): 653-656.

Pommerehne, W. W. and Frey, B. S. (1977). "Public versus Private Production Efficiency in Switzerland: A Theoretical and Empirical Comparison." In *Comparing Urban Service Delivery Systems, Urban Affairs Annual Review,* Ostrom, V. and Pennell Bish, F., eds., Vol. 12, pp. 221-241. Beverly Hills, CA: Sage Publications.

Pryke, R. (1982). "The Comparative Performance of Public and Private Enterprise." *Fiscal Studies* 3 (July): 68-81.

Pucher, J. (1982). "A Decade of Change for Mass Transit." *Transportation Research Record* 858: 48-57.

Pucher, J., Markstedt, A., and Hirscham, I. (1983). "Impacts of Subsidies on the Costs of Urban Public Transport." *Journal of Transport Economics and Policy* 17 (May): 155-172.

Ramanadham, V. V. (1991). *The Economics of Public Enterprise.* New York: Routledge.

Renn, S. C., Schramm, C. J., Watt, J. M., and Derzon, R. A. (1985). "The Effects of Ownership and System Affiliation on the Economic Performance of Hospitals." *Inquiry* 22 (Fall): 219-236.

Rowley, C. K. (1995). "A Public Choice and Social Choice Perspective." In *Economic approaches to organizations and institutions,* Foss, P., ed., pp. 63-104. Hampshire, England: Dartmouth.

Rowley, C. K. and Yarrow, G. K. (1981). "Property Rights, Regulation and Public Enterprise: The Case of the British Steel Industry 1957–1975." *International Review of Law and Economics* 1: 63-96.

Rushing, W. (1974). "Differences in Profit and Nonprofit Organizations: A Study of Effectiveness and Efficiency in General Short-stay Hospitals." *Administrative Science Quarterly* 19 (December): 474-484.

Savas, E. S. (1977). "Policy Analysis for Local Government: Public vs. Private Refuse Collection." *Policy Analysis* 3 (Winter): 49-74.

Schap, D. (1986). *Municipal Ownership in the Electric Utility Industry: A Centennial View.* New York: Praeger.

Schap, D. (1988). "Pricing and Property Rights in Government Enterprises: Some Dynamic Aspects." *Research in Law and Economics* 11: 113-127.

Schlesinger, M. and Dorwart, R. (1984). "Ownership and Mental Health Services: A Reappraisal of the Shift Toward Privately-owned Facilities." *New England Journal of Medicine* 311 (October): 959-965.

Schmitz, Jr., J. A. (1996). "The Role Played by Public Enterprises: How Much Does it Differ Across Countries?" *Federal Reserve Bank of Minneapolis Quarterly* 20 (Spring): 2-15.

Schulz, R. I., Greenley, J. R., and Peterson, R. W. (1984). "Differences in the Direct Costs of Public and Private Acute Inpatient Psychiatric Services." *Inquiry* 21 (Winter): 380-393.

Shepherd, W. G. (1966). "Utility Growth and Profits Under Regulation." In *Utility Regulation: New Directions in Theory and Practice,* Shepherd, W. G. and Gies, T. G., eds., pp. 3-57. New York: Random House.

Shepherd, W. G. (1988). "Public Enterprise: Criteria and Cases." In *The Structure of European Industry,* 2nd ed., de Jong, H. W., ed., pp. 355-388. Netherlands: Kluwer Academic Publishers.

Shepherd, W. G. et al. (1976). *Public Enterprise: Economic Analysis of Theory and Practice.* Lexington, MA: Lexington Books.

Shleifer, A. (1998). "State versus Private Ownership." *Journal of Economic Perspectives* 12 (Fall): 133-150.

Short, R. P. (1984). "Public Enterprises: A Statistical Comparison." In *Public enterprise in mixed economies: Some macroeconomic aspects,* Floyd, R. H., Gray, C. S., and Short, R. P., eds., pp. 110-192. Washington D.C.: International Monetary Fund.

Sloan, F. A. and Vraciu, R. A. (1983). "Investor-owned and Not-for-profit Hospitals: Addressing Some Issues." *Health Affairs* (Spring): 25-37.

Spann, R. M. (1977). "Public versus Private Provision of Governmental Services." In *Budgets and Bureaucrats: The Sources of Government Growth,* Borcherding, T. E., ed., pp. 71-89. Durham, NC: Duke University Press.

Stevens, B. J. (1978). "Scale, Market Structure and the Cost of Refuse Collection." *Review of Economics and Statistics* 60 (August): 438-448.

Tilton, J. E. (March 1973). "The Nature of Firm Ownership and the Adoption of Innovations in the Electric Power Industry." Paper presented at the Public Choice Society, Washington, D.C.

Tuckman, H. P. and Chang, C. F. (1988). "Cost Convergence between For-profit and Not-for-profit Nursing Homes: Does Competition Matter?" *Quarterly Review of Economics and Business* 28 (Winter): 50-65.

Turvey, R. (ed.) (1968). *Public Enterprise.* Baltimore, MD: Penguin.

Vining, A. R. and Boardman, A. E. (1989). "Ownership and Performance in Competitive Environments: A Comparison of the Performance of Private, Mixed, and State-owned Enterprises." *Journal of Law and Economics* 32 (April): 1-33.

Vining, A. R. and Boardman, A. E. (1992). "Ownership versus Competition: Efficiency in Public Enterprise." *Public Choice* 73: 205-239.

Wallis, I. (1985). "The $90m Case for Private Buses." Unpublished paper. Bus and Coach Association, Sydney, Australia.

Wilson, G. W. and Jadlow, J. M. (1982). "Competition, Profit Incentives and Technical Efficiency in the Provision of Nuclear Medicine Services." *Bell Journal of Economics* 13 (Autumn): 472-482.

Zardkoohi, A, and Giroux, G. (1990). "Bureaucratic Behavior and the Choice of Labor input: A Study of Municipal Governments." *Public Choice* 64: 185-190.

Zeckhauser, R. J. and Horn, M. (1989). "The Control and Performance of State-owned Enterprises." In *Privatization and State-owned Enterprises: Lessons from the United States, Great Britain and Canada,* MacAvoy, P. W., Stanbury, W. T, Yarrow, G., and Zeckhauser, R. J., eds., pp. 7-57. Boston: Kluwer Academic Publishers.

Chapter 12

PRIVATIZATION, NATIONALIZATION, AND ASPECTS OF TRANSITION

William S. Peirce
Case Western Reserve University
wsp@po.cwru.edu

Abstract   This chapter treats some of the main themes found in the literature deal-
ing with privatization, nationalization, and transition. It starts by sur-
veying some of the main arguments that have been advanced in favor
of nationalization. Subsequently, it reviews the growth of privatization.
While it might seem axiomatic to some to treat the growth of privatiza-
tion as a manifestation of the triumph of capitalism over socialism, there
are some problematical considerations that would have to be addressed
before any such treatment could be embraced.

1.   INTRODUCTION

Privatization was a popular concept during the last two decades of the 20th
century. Not only did economists write about it, but also governments in many
countries did sell assets to private investors. The most dramatic instances of
privatization were aspects of the transition from communism that accompanied
the collapse of the Soviet Empire, but the phenomenon of governments sell-
ing state owned enterprises (SOEs) could be observed in both developed and
underdeveloped countries throughout the world. Sheshinski and López-Calva
(1998) provide the World Bank estimates that SOE activity as a percentage
of GDP in low income countries decreased from 15 to 3 percent over the pe-
riod 1980 to 1997. While the decreases were less (from a lower base) in the
higher income countries, the ubiquity of the phenomenon is apparent. More-
over, the careful analysis of even core governmental functions to see whether

some aspects could be contracted out or otherwise privatized drew increasing attention.

All of this activity could be interpreted as the triumph of the free market over socialism, which is the theme of Yergin and Stanislaw (1998). Still, a few troublesome questions remain. First, why did the mainstream of economic theory, which only a few years earlier had been so laudatory of nationalization, begin to sing the praises of privatization? Second, how did it happen that policy jumped when theory whistled? Surely this unlagged response is not what we are used to. Third, can the trend toward privatization be monotonic? If so, where did those nationalized firms come from?

The literature dealing with privatization, nationalization, and transition is far too voluminous to summarize here. Instead, some of the main themes will be sketched in the following order. First, some of the main arguments in favor of nationalization will be developed. These will include both the arguments by economists purporting to serve as advisors to a social welfare maximizing government and also some description of reasons asserted for actual nationalizations. The following section reviews some of the changes in economic theory that have tipped the balance toward privatization, still from the viewpoint of the social welfare maximizer, and considers some of the experience with privatization. Section 4 discusses cycles of nationalization and privatization. Section 5 raises the question of whether transition is a separate phenomenon.

2. NATIONALIZATION

2.1. Theoretical Justification

Adam Smith (1776: bk 4, ch.vii, part 3; bk 5, ch.ii, part 1) was not impressed by the quality of management of public enterprises (including the public lands) or, indeed, of other very large enterprises. His reasoning related to the incentives for managers to maximize profits and seek out innovations. After the marginal revolution of the 1870s, however, attention in technical economics turned increasingly to the question of whether the maximization of profits by the firm really led to maximization of social welfare. The answer that perfect competition aligned private and public goals provided scant comfort in an era when firms were forming giant trusts and public utilities and manufacturing firms seemed to enjoy practically unlimited economies of scale (Pigou [1920] (1962): ch. xx-xxii). By the 1930s, the focus on small numbers by Chamberlin (1933) and, especially, product differentiation by Robinson (1933) and the resulting imperfect competition extended the divergence between marginal revenue and marginal cost even to those firms that were small enough to appear competitive to a non-economist.

Once the problem was recognized, the solution was a matter of taste. One might propose allowing the monopolization, but imposing government regulation (Van Hise, 1912). Alternatively, one might propose government ownership (for a case study, see Taylor, 1927). Note that the informal discussion of the incentive and monitoring problems that were so important for Adam Smith were displaced by the greater theoretical virtuosity of welfare economics. Even the attempts of Mises (1935) and Hayek (1935) to discuss information requirements were interpreted by the profession as being satisfied once it was shown that one could imagine the type of information that the socialist planner would have to distribute—without worrying about whether distribution was technically possible or whether anyone would change his behavior in response to the information.

By the time that the British Labour government had its chance to nationalize industries following World War II, the stated grounds for nationalization had expanded (Robson, 1962, pp. 29-45). Coal mining was included because the industry was in such wretched condition. Steel was a target because it was a healthy and essential industry. Moreover, Labour party doctrine called for ownership of the "Commanding Heights." That is, enough basic industries to enhance the possibility of control over the economy (Robson, 1962, p. 42; Yergin and Stanislaw). Nationalization, it was supposed, would also lead to harmonious labor relations, an end to the wastes of competition, and a decreased emphasis on the profit motive.

2.2. Nationalization in Practice

Before an enterprise or activity can be privatized it must be in government hands. Where did these government activities come from? Why were they formed by government or nationalized? One practical reason for the state to operate an enterprise is to obtain revenue. Study of SOEs was a part of the public finance course until it was displaced by Keynesian macroeconomics. The public finance text by the prominent Princeton economist Harley Lutz (2nd ed., 1930), for example, devotes Part I to Public Expenditures and Part II (pp. 135-257) to "Public Revenue except Taxation." The bulk of that deals with public land and industries,

Revenues from the public domain varied by country and time, depending on the way in which the feudal structure evolved into the modern state with separation between the budgets of the state and the ruler. Kindleberger (1984, p. 173) mentions that Frederick I (the Great Elector) of Prussia in the 17th century had at his disposal both taxes and the yield of the royal domain. The House of Hollenzollern owned land containing one-quarter to one-third of the peasants in the country. The yield included rents, profits on production for market, tolls, taxes, and revenue from monopolies. For countries following Roman

law, the sovereign retained the mineral rights, regardless of the ownership of the surface, and hence was entitled to royalties, but often operated the mines, as well.

Backhaus and Wagner (1987) cite numerous estimates of the fraction of the government budget generated by government enterprises. The numbers range widely by time and place, but exceeded half of total revenues in Prussia and Saxony at the end of the 19th century.

State monopolies have been traditional sources of revenue in a number of countries. Lutz (1930) lists such government monopolies as matches and gunpowder (France), salt and life insurance (Italy), alcohol (Sweden and Russia prior to 1914), and camphor (Japan). Madsen (1916) focuses on the tobacco monopolies of France (established by Colbert in 1674), Italy (taken over from the states in 1861), Austria (instituted in 1670), and Japan (began operations in 1898). He mentions briefly several other tobacco monopolies, as well as other monopolies, including "The Indian Opium Monopoly which may be said to exist, not for the sake of revenue, but as a means to prevent the excessive consumption of a harmful drug" (Madsen, 1916, p. 10). Presumably one could depress consumption by raising the price above the profit maximizing point.

Madsen (1916, pp. 10-11) notes that the revenue generating monopolies "… are more often than not descendants of the old-time monopolies and patents granted by kings to their favorites, or they are an out-growth of the much-discredited device of farming the revenues, the government monopoly having become a directly administered State affair instead of a joint concern with the state and privileged individuals as partners."

This is the topic that Congleton and Lee (2000) deal with. Presumably, a government seeking revenue could set a stiff excise tax on the products of a competitive industry, monopolize the industry under state ownership, or sell or lease the monopoly privilege to a private party. Not only does the sale of the monopoly privilege (or other forms of tax farming) reduce the administrative burdens on the government, it also provides immediate revenue.

If the government has a rate of time preference that exceeds the discount rate in the private sector, the price at which it can sell the right to collect monopoly profits in the future is worth more than the value to the government. Selling assets to balance operating budgets is usually a route to financial ruin, however. (It is interesting that reducing government deficits has been one of the strongest motives in the current wave of privatizations.) Governments, of course, control tariffs and internal regulations that influence the annual revenues and durability of the monopolies granted, and thus the credibility of the government becomes an issue. As the government gets shakier, its rate of time preference increases, but so does the uncertainty associated with its promises.

The state enterprise established for the purpose of generating revenue should enter the very markets where excise taxes prove most profitable. These are the widely consumed commodities with low elasticity of demand, such as salt, matches, tobacco, and gunpowder. Whether the state chooses to tax the poor, or to exploit them by operating a monopoly, or to auction off the right to exploit them is a matter of expediency in the particular circumstances.

Profits are not the only reason adduced for state ownership. Losses are an equally prominent reason. When important industries, especially those employing a high percentage of the labor force in particular locations, seem unable to continue under private ownership, the political pressure for state intervention is strong. Sometimes the state provides or guarantees loans and eventually forecloses. At other times the acquisition is more direct. Shipbuilding, steelmaking, and mining were common examples in the decades following World War II, but no industry was excluded from this inadvertent nationalization.

Italy carried this to an extreme. According to Rossi (1955), businessmen commonly set up banks to finance their own firms with deposits from small savers. The banks would be very profitable during booms, but then would be threatened with failure during depressions. The government took over banks to protect the depositors, and thus acquired interests in a wide assortment of businesses. This pattern began around 1900 and did not really change after the Mussolini government took office in 1922. In 1933 the Institute of Industrial Reconstruction (IRI) became the successor to the various agencies dealing with the bank failures. Martinelli (1981) recounts how the IRI and similar state agencies continued to accumulate partial or total ownership of a bewildering array of firms until the privatization wave overtook Italy at the end of the 20th century.

War and its aftermath have also been sources of state owned enterprises. Under the controlled, forced draft conditions of war, it is not surprising that munitions industries, shipyards, powerplants, and even the related housing stock are government owned. Even when government is not the owner, it may have guaranteed loans for construction or expansion, which may lead to government ownership in the recession following the war.

The dislocations of war act in other ways, as well. Following World War II, for example, Austria was left with a steel industry that had been restructured under German ownership during the war. According to Rosegger (1985), government ownership seemed to offer an easy solution to the tangled claims on ownership. In France, property of some collaborators was confiscated after World War II. Since nationalization was a rallying cry of the French Resistance (Byé, 1955, p. 74), it is not surprising that, for example, the automobile company of Louis Renault was nationalized. Press nationalizations provoked the

additional explanation that they were "… to prevent rotten journalism from returning in its old form or in some new guise" (Byé, 1955, p. 76n).

Revenge can be a powerful motive, even in the absence of wartime provocations. Chua (1995) builds a theory of nationalization-privatization cycles (discussed below) around the envy of the impoverished masses over the wealth of an identifiable minority group in certain Latin American and Asian countries.

Control of individual behavior is another justification for state ownership. This is one ostensible reason for state ownership of lotteries, liquor stores, and methadone clinics, as well as the Indian opium monopoly. The revenue-generating objective is sometimes at odds with the control objective, as, for example, when a state lottery advertises.

The macro aspect of control inspired much of the rhetoric surrounding the post World War II nationalizations (Robson, 1962, pp. 29-45). The central bank must be nationalized to ensure monetary control. Large firms in key industries must be nationalized to give the government instruments to control the remaining private sector. Key export industries and major firms on the technological frontiers had to be nationalized to ensure that the nation was not left behind in chemicals, electronics, nuclear power, aircraft production, computers or whatever happened to capture the political imagination at the moment.

Often social objectives were explicitly introduced. The nationalized coal mines should be safer and provide better wages and working conditions. State owned utilities should provide households with low rates, protect the environment, and foster economic development.

Once social objectives enter the objective function, even if the objective is economic development, the nationalized firm is freed from the objective criterion of generating profits. While it is possible to argue that, with enough information, politicians could determine whether the profits forgone have purchased enough of the additional desiderata, the technical and information requirements for doing the analysis are substantial (Backhaus, 1994).

2.3.   Form of Organization

Efforts at nationalization are immediately confronted with an organizational choice, or perhaps an organizational dilemma. If the nationalized activity is established within an ordinary government agency, it becomes subject to all of the usual much maligned bureaucratic red tape. This is a very inefficient environment in which to conduct a business. If, on the other hand, the activity is set up as a government-owned corporation, a different set of problems arises.

In the large, well organized, western democracies government procedures are structured with the view to control two types of corruption. The first, against which private firms also build defenses, is private enrichment by employees at the expense of the organization's objectives. This includes the routine thefts of cash, supplies, and equipment, as well as more complex transfers

of assets, sales to the organization at excessive cost, kickbacks on purchases, bribes, etc.

The second, and more distinctive, form of corruption in the public sector of a democracy is use of administrative agencies by the ruling party to maintain its power. That is, can the tax authorities be induced to use selective audits as a device to harass critics of the party in power? Will the FBI provide files on prominent members of the opposition? Will export restrictions be waved for large contributors? Can various forms of public funding be shifted to reward supportive firms and localities?

The civil service employees are supposed to be professional enough to resist such improper requests. That implies that they must be able to refuse orders without being fired, demoted, or transferred to the other end of the country. Such people will be difficult to manage. Indeed, Hart and Moore (1990) stress that ownership of a firm means the right to fire selectively—to get rid of individual employees. By this standard, no one owns government agencies.

Moreover, the output of the traditional government agencies is difficult to measure, so control systems have evolved that provide detailed measurement and control of inputs and procedures along with detailed personnel rules and effective lifetime security of employment. The reason that the bureaucratic procedures of government are so resistant to reform is that they are absolutely essential for the integrity of the political process and the security of the professionals who produce unmeasured outputs in a politically charged environment (Warwick, 1975).

Nationalized enterprises produce at least some output that can be valued and purchased by paying customers. Thus, much of the bureaucratic red tape can be replaced by the standard control devices used in private firms, including measurement of output and costs, rewarding those who perform, and firing those who do not. It is very difficult, however, to insert a small island of performance-based management in a sea of bureaucratic procedures. The key problem is that the organization and sanctioning of such an enclave would have to be entrusted to the traditional administrative structure.

These are old problems, and the old answer was thought to be the government owned corporation. Van Dorn (1926) studied the World War I era corporations established by the U.S. government. These included, among others, the Emergency Fleet Corporation, the United States Grain Corporation, the United States Sugar Equalization Board (which used its monopsony of Cuban sugar to subsidize consumer prices for all sugar), and the United States Spruce Production Corporation (charged with ensuring the supply of the material essential to aircraft production). The corporate form did provide strong administrators with the flexibility to purchase, hire, and sign long term contracts without being delayed and encumbered by standard bureaucratic rules. Typically, one particular cabinet level official would be authorized to hold and vote the shares

of a corporation, organized in the usual form, with initial capitalization appropriated by Congress. In the case of the U.S. Grain Corporation, the President appointed four of the directors, who then chose the remaining three. Thus the organizations were entirely bereft of control except by the President and an occasional congressional inquiry. At best, this would result in an extreme form of managerial control. It could, however, permit fairly crude distributional politics, which apparently occurred when the Spruce Corporation concluded its operations (Van Dorn, 1926, pp. 253-254).

The reason for belaboring these points is that the issues recur. Rosegger (1985) notes that during the great wave of Austrian nationalization after World War II the organizational forms of the firms were maintained intact. The government became the sole shareholder. The Federal Chancellor thus attended an annual meeting, for each of the dozens of corporations, and cast his vote for the board of directors. Any formal attempts to coordinate the actions of separate state corporations for purposes of industrial policy or to exert special pressure on state owned firms to behave differently from other managerially controlled firms were essentially impossible.

Robson (1962, ch. III) discussed the public corporation in Britain. Although a number of variations persisted, in general the formality of shareholders and the annual meeting did not exist. The general principle had been established that while Parliament could guide policy, it could not interfere in administration. Indeed, Normanton (1981) pointed out that the British nationalized industries were not even subject to state audit.

Since, in the general British case, surplus (profits) could not be paid out to the owner or to management, it had to be accumulated as a reserve, reinvested in the business, or devoted to improving service or working conditions. Alternatively, prices could be reduced. Thus it would seem like the ultimate in managerial control. Yet each government corporation existed within the jurisdiction of a cabinet minister and hence was subject to immediate political influence, as well as the general policy directives from Parliament. Moreover, the chief executive of a government corporation is frequently someone in the midst of a political career or with political aspirations; hence political influences can be direct. The government corporation escapes the red tape that entangles government agencies, but one corollary is the greater exposure to political pressures (Grassini, 1981; Anastassopoulos, 1981) with consequences detailed by Shleifer and Vishny (1994).

3.   PRIVATIZATION

3.1.   Progress in Economic Theory

The most optimistic explanation for the advance of privatization is that economic theory has demonstrated the superiority of private ownership (under

certain conditions) and the world is rushing to comply. In their very fine re-view of the economic theory and the empirical record, Sheshinski and López-Calva (1998) convey this flavor without actually stating it. Yergin and Stanis-law (1998) were not so restrained. While an economist may be attracted to this position, it is not fully convincing. After all, the free trade arguments of Smith and his successors of the past 200 years are not really believed by politicians even now when they are forced to comply with reduction of trade barriers. More to the point, the generation of economists who promoted nationaliza-tion in Great Britain had been raised on Adam Smith's pungent remarks about monopoly and especially government monopoly. The past couple of decades have seen advances in the microeconomics of incentives and contracting, but the fundamental works of Hayek (1935,1944) on prices and socialism that are widely cited now were available half a century ago.

The main reason for doubting that economic theory led the way, however, is that the recent wave of privatization seemed to begin with the Thatcher govern-ment, elected in 1979. (Chile had seemed to be a curious exception to the usual way of doing things when it began its privatization program in 1970.) While some economists could always be found who would argue the case for priva-tization, the major advances in economic theory that appeared to justify that view to the mainstream of the profession appeared after the political change had begun.

While the economic analyses that seemed to support nationalization had stressed characteristics of technology—economies of scale and externalities of production—the analyses that cast doubt on the benefits of nationalization drew on the developing economic literature in industrial organization relating to such concepts as control, contracts, incentives, and monitoring. It is difficult to find any elements in this literature that were not incisively developed by Tullock (1965), and indeed many of the elements were common in the older literature of management and public administration reviewed by Peirce (1981), but the more formal literature gave these old ideas new rigor and moved them into the economic mainstream.

Most of the recent economic theory arises out of work on the relationship between owners and managers or between the government control agency (e.g., a public utilities commission) and the manager. The economic theory need not be so narrowly restricted, but very few of the implications of moving to a different level of analysis have been worked out. At the more aggregate level, one could inquire about the cost to the government of taking on one new program. A traditional cost-benefit analysis implicitly assumes that monitor-ing by the legislature, the chief executive, and the voters is costless and that interactions with other parts of government will absorb no resources.

At the less aggregated level, one could ask whether the manager of the na-tionalized firm has the same level of control over his subordinates as does the

manager of the private firm. Factors influencing this include the usual salary compression in the SOE, the inability to use ownership of shares as an incentive, and possible political constraints on layoffs, either generally or of specific people. Moreover, public enterprises ordinarily have multiple objectives. If one were concerned solely with profit maximization, there would be little reason to nationalize the firm. As Raiffa (1981) suggested, the manager will have difficulty in monitoring the performance of his subordinates when the political environment makes it impossible to state precise tradeoffs among a long list of objectives. [Arrow (1981), however, did not see that as a problem.]

The remainder of this discussion will focus on the level of aggregation that is typical of modern economic theory. In this brief summary of the issues, no attempt is made to duplicate the excellent review by Vickers and Yarrow (1988). One crucial insight of the literature is that agents do not always act in the interests of their principals. The principal can write a contract that provides positive and negative sanctions for particular courses of action. Contracts can never be complete, however, because it is impossible to know all possible future states of the world. Moreover, the principal must monitor performance of his agents, and perfect monitoring would require an infinite amount of the principal's time.

More fundamental questions concern the identity and interests of the principal. In a liberal democracy, in contrast to autocracy, presumably the interests of the people are paramount. Most economists are trained to think in terms of arrangements that maximize social welfare, defined in terms of individual preferences. The concept of a social welfare function is technically meaningless, of course, because of the inadmissability of interpersonal welfare comparisons and the lack of any operational measures of utility (Little, 1957). Thus, practical advisors to government usually end up by adopting the standard of maximizing gross domestic product (Harberger, 1971), whatever their misgivings about the conceptual weakness of such an objective.

Even in democracies, however, the people do not serve as the principal—governments do. Government, including the monitoring of nationalized industries, is a political process. Objectives other than maximizing social welfare or gross domestic product motivate the participants; and the most plausible motive to assume is self interest.

This brings us full circle because the virtue of private ownership is, under the right conditions, to maximize national income. The question now becomes an empirical one; i.e., if we know that the conditions for perfect efficiency are not met because of such matters as monopoly, economies of scale, externalities, and separation of ownership and control can public ownership, with its own characteristics including imperfect monitoring by self-interested politicians, improve the situation. Such empirical work must pay careful attention to questions of control, as well as ownership, because the realistic alternative

to public ownership is often heavily regulated private ownership. Rowley and Yarrow (1981) in a careful example of such empirical research found, in addition to the politically constrained location of major investments, a failure by the British steel industry in its nationalized period to adopt modern technology rapidly.

From a theoretical perspective the failure to innovate may be surprising. If one thinks of innovation as risk-taking, the public manager bears less risk of failure than does the manager of a private firm, who may lose his job during bankruptcy proceedings or even if profits do not meet expectations. Governments rarely impose binding budget constraints on firms that they own. The consequence of a deficit is a larger subsidy. Although the risks of failure are less in the nationalized setting, the rewards for the managerial burdens of a major innovation are also less, and may even be negative if employment suffers or the location of economic activity changes.

3.2.   Reasons for Privatizing

Increasing dissatisfaction with the poor performance of many public enterprises in the difficult economic environment of the 1970s certainly contributed to an atmosphere conducive to privatization. Bernholz (2000) cites "... international competition among states …" as a force that motivates political leaders as diverse as Thatcher, Deng, and Gorbatchev. The nationalized industries, by pursuing noneconomic objectives, were dragging down the economic performance of many countries.

More concretely, the wave of privatization in the 1980s was given a major impetus by budgetary problems. Despite the historical role of state owned enterprises as a source of revenue, and despite the discussion of the social objectives that could be pursued with the surpluses that such enterprises were supposed to generate, the more typical situation was massive deficits that drained state treasuries and constrained the politicians from providing more popular programs or tax cuts. The commitment to budgetary balance was reinforced by the International Monetary Fund, which made deficit reduction a condition for its loans. For such purposes, countries could eliminate an annual expenditure for subsidizing the losses of the enterprise, and simultaneously count the one-time receipt from sale of the asset as though it were revenue in the annual budget. Countries striving to meet the Maastricht Criteria for joining the European Monetary Union in the 1990s followed the same procedure.

3.3.   Privatization of Governmental Functions

The discussion to this point has dealt with the private or state ownership of firms that provide goods or services for sale in private markets. The term "privatization" has, however, been used in another sense; that is, the transfer

of certain aspects of public functions to the private sector. For example, the payroll checks for government employees might be prepared by a private contractor. The boundary between the two uses of the word is unclear, depending in part on what is considered an governmental function at a particular time and place.

Privatization can take a variety of forms. The spectrum of possibilities ranges from such modest changes as contracting out minor aspects of some government function (e.g., housekeeping services or data entry) to the extreme of abolishing some agency and its functions (e.g., the CAB and the ICC). In the middle of the spectrum are such possibilities as government contracting for public services, such as garbage collection, and taxpayer-funded vouchers that allow subsidization of particular private goods with reduced government control (e.g., food stamps or school vouchers). Data showing the growing importance of such forms of privatization are difficult to obtain on a consistent basis, but casual observation suggests that such devices are becoming more important. They are certainly being discussed more.

What accounts for this privatization movement? Why is it occurring now? Most important, will it persist? Savas (1982, 1999) summarizes much of the theory and offers numerous case studies. Government involvement in an activity; e.g., primary education, can range along a continuum from zero (let families and private organizations take care of it) to compulsory attendance for all children in assigned schools. Intermediate points can include choice within some set of government schools, the option to use unsubsidized private schools, and vouchers to subsidize some or all of the costs of private schools. The schooling case generates heated discussion of fairness, indoctrination, and quality. The issues might be simpler if the discussion were confined to whether government schools should operate their own buses and cafeterias or contract out those ancillary services.

Mueller (1989, Table 14.1) summarizes numerous empirical studies and finds that most indicate savings from privatization. It must be noted, however, that the interface between the private firm and the government agency is always sensitive and subject to corruption in the form of bribery or political influence. A private monopoly, especially with entry blocked by politicians, might be more expensive than production of the ancillary service by government employees. Sometimes contracts are complex (but incomplete) and physical assets specialized, so holdup problems can occur. The political environnment is a tricky one in which to renegotiate contracts. Indeed, it is often not clear what the incentives are for the government negotiator.

It is clear that competition is essential to force costs down, regardless of ownership. Indeed Vickers and Yarrow (1988) note that competition seems to be more important than ownership as a determinant of efficiency. Assuming that to be the case, then the movements in the 1980s and 1990s to liberalize

economies, in the sense of reducing internal and external barriers to competition, would have been expected to increase the efficiency of various activities, regardless of ownership. That raises the interesting question, not considered here, of the political interaction between government ownership and liberalization.

In the context of the make-or-buy decision of the government agency, crucial issues include the incentives of the people who determine what will be contracted out, the structuring of contracts and bidding to include or exclude competition, and the monitoring of performance. Government contracting can be as subject to rent-seeking activities as direct government production. It may also be subject to cumbersome procedures adopted in the name of reducing political influence, but sometimes serving to reduce entry.

A way around this problem is to delegate the decisionmaking and monitoring functions to the public, assuming this is to remain a government activity at all. That is the power of vouchers and similar programs (food stamps) from a social welfare maximizing view. The consumer makes the decisions and has the usual incentives to consider marginal costs and marginal benefits, if the program is structured correctly. But it does raise a question from the perspective of the politician: What is the reward in fund-raising or vote-getting or in personal perquisites for adopting such techniques?

Many government functions are not amenable to meaningful monitoring by the public. Hart, Shleifer, and Vishny (1997) are dubious about privatization of prisons because the strong private incentive for cost reduction may dominate the less easily monitored aspects of quality; e.g., a safe and humane environment. Since the authors assume, in this model, that cost reducing innovations decrease the quality of service, the conclusions will not convince those who doubt that public prisons are efficiently operated.

The more general problem is that prisoners, in Hirschman's (1970) terms, have a weak "voice" and are not supposed to "exit," so they are not in a good position to effect changes when they observe lapses in quality. Voters generally do not think much about the issue. A very strong voice, however, is that of the public employees union. Indeed, López-de-Silanes, Shleifer, and Vishny (1997) find that strong unions of public employees are associated with lower rates of privatization.

4.   NATIONALIZATION—PRIVATIZATION CYCLES

Nationalization-privatization cycles (NPCs) have received little explicit attention despite the obvious fact that governments could not privatize activities today if they had not previously acquired them. Siegmund (1997) finds NPCs in the sense of transitions from nationalization to privatization and back, but without regular periodicity. Average phase lengths differ in different countries.

One feature common to all countries studied is that the beginning and the end of the 20th century were periods of privatization.

The few published explanations for NPCs differ widely. At one extreme, Rosa (1993) postulates a unitary, optimizing government that buys and sells its enterprises as fluctuations in the discount rate and the marginal social cost of tax revenue make the enterprise worth more or less to the government than to the private sector. At the other extreme, Kingston (2000) writes of cycles that may last centuries in which the efficiency of unconstrained property rights gives the owners of property increasing political power leading to excess, decadence, and revolution.

4.1.   The Cycle as a Political Process

The remaining discussion of cycles will focus on the middle ground where self-interested people in politics and other walks of life contend for rents under whatever circumstances happen to prevail. In the most substantial such public choice analysis, Backhaus (1989) notes that the most important difference between public and private ownership of an enterprise is the limited duration of ownership rights in the public sector. As soon as the party in power loses an election (or otherwise loses political control), the benefits of ownership of the SOE are passed on to the adversary. The profit-generating capacity of the SOE can be used for providing good jobs to reward party faithful, for discretionary spending on programs that are expected to increase popular support (or incomes of politicians and supporters), and for net income that can be returned to the state treasury to support unrelated programs that increase political popularity. Because the short time horizon reduces the returns that politicians can expect from the job seekers and the beneficiaries of discretionary programs, the incumbent politicians try to extend civil service (or similar) tenure protection to the SOE and to pass legislation requiring the continuance of favored discretionary programs. When a new government takes office, old programs and people are hard to dislodge (Moe, 1997, p. 466), so new programs and offices are grafted on. After a few changes in government, the net income available to the treasury will be depleted. Indeed, the SOE may require a subsidy from general tax revenue.

Compounding the problem in the basic story is the effect of the short time horizon on investment. Backhaus notes that investing profits now to yield benefits later offers little appeal. The politician could do something that would increase political popularity now with that profit. Moreover, much of the payoff from an investment will accrue after the next election, which means that the opposition might have control over it. One might add that such investment as occurs might be guided by political considerations. Even if the administrative laws and traditions of the country are strong enough to prevent outright corruption, the location, and sometimes the technology, of the investment may be

strongly influenced by the political power of particular regions, the imperative to use domestic technology, or other non-economic considerations.

In the effort to prevent the next government from undoing particular discretionary projects and removing particular personnel, the SOE may be burdened with regulations that make control of the labor force extremely difficult to achieve. Insulating the bureaucracy from the political pressure of the next administration also attenuates managerial control.

At this stage, the SOE becomes a good candidate for privatization. It has become a net drain on the treasury, its available resources are legally committed to uninteresting programs, the existing personnel resist redirection of their efforts, the capital stock is inadequate and inappropriate, and it is so overstaffed that supporters can no longer be placed in jobs there. The privatization negotiations will center around the degree of protection given to the labor force and the necessity of maintaining particular services, as well as the price. (Clarke and Cull (2000) discuss the provisions of some privatization contracts.) In an honestly administered country, the incumbent party will control use of the revenue from the sale of the SOE to provide services that increase its popularity. In a corrupt country, politicians may find some way to increase personal wealth directly.

Dániel (1997) provides details of the cycle in the Hungarian housing market. The local government owns housing, charges low rents, loses money, and maintains the housing stock poorly. The losses accumulate to the point where the revenue from the sale of the buildings is attractive to the party in power. The buildings are sold cheaply as condominiums, often to the retired elite who live in the higher quality apartments. The big losers are young people who cannot afford downpayments. Thus the stage is set for the next round of government intervention.

4.2.   The Chua Study

The wonderfully detailed study by Chua (1995) begins with statements of politicians using the same words—modernization, democracy, justice, and efficiency—to advocate either privatization or nationalization, depending on the time and place. Chua studied a number of postcolonial Latin American and Asian countries characterized by sharp ethnic and economic divisions. She noted that privatization, as implemented, "… resulted in the disproportionate prosperity of particular, ethnically identifiable groups" and "ethnically charged... nationalist movements have repeatedly succeeded in overturning regimes championing private enterprise" (p. 226). The evidence is ample and convincing that NPCs have occurred as she describes them in a number of countries including Mexico, Brazil, Chile, Peru, Argentina, Malaysia, and various other Asian countries.

Does Chua's model constitute a supplement to, or a refutation of, the model developed by Backhaus, which is informed by the European and U.S. experience? Certainly the story of the growing burdens placed on the SOEs is common to the two models. The tendency for government to resemble a stationary bandit (Olson, 2000) is often less disguised in the underdeveloped nations than in those with a longer indigenous democratic tradition. Thus, the nationalized industries may be vehicles for overt corruption, rather than the more regular and legal transfers in the form of wage rates exceeding market rates, cross-subsidization of favored customers, and special treatment of favored suppliers. But surely these are differences in degree, not in kind.

Both privatization and nationalization can be used to provide favors to specific individuals. If, for example, the government buys a firm that is slipping into bankruptcy, is it to help the Prime Minister's cousin, who owns it, or to save the jobs of those who work there? Does it matter? At the other change of phase, the Russian experience has demonstrated that privatization can be synonymous with theft of state assets. Presumably it is easier for the taxpayers to capture a significant share of the value when capital markets are functioning fairly well, the appropriate auditing institutions are in place, and a well organized opposition is watching.

The use of nationalization to move highly visible and symbolic firms out of the hands of foreigners or hated ethnic groups is apparently a distinctive feature of the cases that Chua studied. Yet is it really so different from the efforts of French governments to build Machines Bull into a force that could repel the Alien Invader—IBM? Moreover, in the cases Chua cites it would be interesting to know whether the "victims" of nationalization suffered financially. Since the same minorities seem to reappear in the next round, maybe the wild rhetoric is just a convenient cover for the underlying processes described by Backhaus.

Recognition of NPCs as a political process sheds some light on issues that have arisen during the recent privatizations. For example, governments are usually said to have great political difficulties in laying off underemployed workers. Yet it is common for governments to "clean up" a nationalized firm by laying off excess labor and assuming pension or other obligations before selling the firm to private investors (Djankov and Pohl, 1998).

Often the sales are made below the market valuations, as indicated by the subsequent immediate increase in share prices (Vickers and Yarrow, 1988, pp. 173-181). This is quite straightforward if it is recognized that the employees of the nationalized firm were not going to vote for the privatizing party anyway, but those who buy shares at a favorable price will be highly receptive to further privatization efforts and will resist renationalization.

The question of whether first to privatize monopolies or first to liberalize the economy to destroy the barriers that strengthen the firm's monopoly power

is answered differently depending on one's approach. The barriers are most likely to be removed if that is done first. The government can extract more revenue from the sale of a protected monopoly, however, and once that monopoly has been sold to the private sector, any attempt to remove barriers will encounter formidable opposition from those who have just purchased it.

5.   TRANSITION

The question then arises whether transition, in the recent sense of the change from communism to a market economy, can be subsumed under the analysis that has been developed for NPCs. Transition, of course, encompasses a wide range of changes in politics and culture, as well as in the economy. This brief discussion will not attempt to deal with any aspect except the change in the ownership and control of firms, and even within that restricted field will focus primarily on a few aspects of the Russian experience.

In order to analyze the transition, it is useful to start from some model of the functioning of the Soviet economy prior to 1989. The approach taken here is that of Anderson and Boettke (1997) and Olson (2000). Although there are differences of emphasis between these two interpretations, they are both based on the assumption that the various participants in the Soviet economy pursued their own interests. This contrasts with the standard socialist model that assumes a planning apparatus dedicated to the maximization of social welfare.

The Olson model analyzes the Stalinist economy as a cleverly designed engine for maximizing the ruler's command over resources. These were used to buy security for the ruler against the threat of domestic insurrection and foreign invasion. Over time, the system deteriorated for two reasons. First, as is widely recognized, the central planning process cannot deal effectively with technological change. Second, the growing strength of groups with such narrow interests that they had no interest in the size of the economy—only in their own shares—inhibits growth, according to the analysis Olson developed in his earlier books (1965, 1982).

Anderson and Boettke (1997) deny that the Soviet economy was centrally planned. They analyze it as a typical example of mercantilism. That is, various participants controlled monopolies that were protected by internal and external regulations. Despite the multitude of quotas and other symbols of physical planning, Anderson and Boettke argue that participants in the economy were still reacting to the signals provided by prices, even though many of the transactions were illegal. The autocrat in a mercantilist economy buys support from powerful allies by putting them in positions where they can extract some share of the autocrat's revenues. The Soviet central planning structure was a mechanism to protect the turf of the monopolies.

Under either version of this theory, the Soviet economy became increasingly inefficient over time, but the incentives for individuals were structured in such

a way that it did not pay anyone to try to do anything other than pursue his own rents until the inefficiencies led to collapse.

More significant for the transition, however, was the discrepancy between the legal institutions of property and the established practice. When a monarch sells a monopoly privilege, as Congleton and Lee (2000) discuss, the buyer has the legal title to the physical assets of the firm he establishes under that royal grant. If a tax is farmed out, however, the recipient of that revenue stream has no asset if there is a change in regime. The Soviet case, legally seemed to resemble the latter one for all participants. Rents were attached only to position, not to ownership of property.

Once the central controls decayed to the point where change appeared inevitable, some, but not all, officials were in a good position to acquire ownership of state assets (Blasi, 1997). Much of this informal privatization occurred before the official progams (Shleifer and Treisman, 2000). Privatization progressed so rapidly that Boycko, Shleifer, and Vishny (1993, p. 139) declared that, "… privatization has become the most successful reform in Russia." However, some powerful officials in the old regime who had rights only to income streams (bribes for permits, for example) or power (the military and the KGB) were not in a position to acquire property and thus lost relative, and sometimes absolute, status and income. Some became opponents of the new "oligarchs" who had been in favored positions to seize the best assets. Those who acquired assets became opponents of further liberalization that might threaten the value of their monopolies (Shleifer and Treisman, 2000, p. 12).

The most valuable of the state assets were those based on natural resources, where the value depends on world markets. Much of the industrial machinery, by contrast, would be worth nothing in an open and free economy. Hence, the new "owners" of such assets become a force for maintaining the old tangle of monopoly privileges and regulations.

The initial scramble for ownership and later efforts to retain control were abetted by the property rights anarchy in Russia that resulted from the many decades of official prohibition of private ownership. If the only problem had been the absence of certain laws, any of the advanced industrial economies could have supplied documents to be translated and adopted. Writing laws, however, is part of the political process that can involve major shifts in power and wealth (Gustafson, 1999, p. 154; Rapaczynski, 1996).

Some have looked at the great fortunes of the oligarchs and remarked on the apparently massive redistribution of wealth. Gustafson (1999, p. 26) suggests that "Never in human history, perhaps, has there been such a dramatic and sudden transfer of wealth, other than through military conquest." Undoubtedly one can find instances of the clever and energetic boy from the provinces who made a fortune from the privatization program. But such a lad probably would

have done well under the old regime, as well. More to the point, it appears according to anecdotes, scraps of empirical data, and formal modeling (Alexeev, 1999) that the main effect of Russian privatization was not to change the rankings in the distribution of wealth, but to magnify the initial inequality.

One could argue that the initial distribution of wealth will be largely irrelevant, as long as Russia adopts the liberal policies (free trade externally and internally, security of person and property, etc.) that are conducive to entrepreneurship and a market system. It is not clear where the support for liberal reforms will originate. To the extent that mass privatization succeeds in giving voters a stake in private property, Schmidt (2000) argues that the risk of expropriation is reduced. Yet to the extent that people see both their jobs and their financial wealth tied to a particular firm, the incentives to maintain the structure of regulation that keeps that firm profitable are strengthened. Most important is the question of whether the distribution of wealth is tenable given the existing distribution of political power.

6.   CONCLUSION

The big questions raised in the Introduction—Why did theory change? Did policy respond to theory? and Is privatization a monotonic process?— can now be answered tentatively. The usual historical pattern can be observed in which political events force the mainstream of economic theory to change course. Keynes had the direction of causality backwards. Privatization is not a monotonic process. It is one phase of a continuing political process, which may, as Buchanan (1997) suggested, lead to mercantilism via regulation before it leads back to nationalization.

REFERENCES

Alexeev, M. (1999). "The Effect of Privatization on Wealth Distribution in Russia." *Economics of Transition 7* (2): 449-465.

Aharoni, Y. (1981). "Managerial Discretion." *Vernon and Aharoni* 13: 184-193.

Anastassopoulos, J. (1981). "The French Experience: Conflicts with Government." *Vernon and Aharoni* 7: 99-116.

Anderson, G. and Boettke, P. (1997). "Soviet Venality: A Rent-seeking Model of the Communist State." *Public Choice* 93 (1-2): 37-53.

Arrow, K. (1981). "On Finance and Decision Making." *Vernon and Aharoni* 4: 63-69.

Backhaus, J. (1989). "Privatization and Nationalization: A Suggested Approach." *Annals of Public and Cooperative Economics* 60 (3): 307-328.

Backhaus, J. (1994). "Assessing the Performance of Public Enterprises: A Public Choice Approach." *European Journal of Law and Economics* 1 (4, December): 275-287.

Backhaus, J. and Wagner, R. E. (1987). "The Cameralists: A Public Choice Perspective." *Public Choice* 53 (1): 1-20.

Bernholz, P. (2000). "Democracy and Capitalism: Are They Compatible in the Long-run?" *Journal of Evolutionary Economics* 10 (1-2): 3-16.

Blasi, J., Kroumova, M., and Kruse, D. (1997). *Kremlin Capitalism: The Privatization of the Russian Economy.* Ithaca, NY: ILR Press of Cornell University Press.

Boycko, M., Shleifer, A., and Vishny, R. (1993). "Privatizing Russia." *Brookings Papers on Economic Activity* 2: 139-192.

Buchanan, J. (1997). "Socialism is Dead but Leviathan Lives On." Ch. 10 in *Post-socialist Political Economy: Selected Essays,* pp. 83-92. Cheltenham, UK and Lyme, US: Edward Elgar.

Byé, M. (1955). "Nationalization in France." In Einaudi et al. pp. 65-188.

Chamberlin, E. [1933]. *The Theory of Monopolistic Competition: A Re-orientation of the Theory of Value.* Cambridge, MA: Harvard University Press, 2nd ed., 1936.

Chua, Amy L. (1995). "The Privatization-nationalization Cycle: The Link Between Markets and Ethnicity in Developing Countries." *Columbia Law Review* 95 (2): 223-303.

Clarke, O. R. G. and Cull, R. (2000). "Provincial Bank Privatization in Argentina: The Why, the How, and the so What." Paper presented at the *Public Choice Society Meetings.* Charleston, S.C., March 10-12, 2000.

Congleton, R. and Lee, S. (2000). "Endogenous Rent-seeking: Schumpeter and Ramsay meet Tullock and Buchanan." Paper presented at the *Public Choice Society Meetings.* Charleston, S.C., March 10-12, 2000.

Dániel, Z. (1997). "The Paradox in the Privatization of Hungary's Public Housing: A National Gift or a Bad Bargain?" *Economics of Transition* 5 (1): 147-170.

Djankov, S. and Pohl, G. (1998). "The Restructuring of Large Firms in the Slovak Republic." *Economics of Transition* 6 (1): 67-85.

Einaudi, M., Byé M., and Rossi, E. (1955). *Nationalization in France and Italy.* Ithaca, NY: Cornell University Press.

Grassini, F. (1981). "The Italian Enterprises: The Political Constraints." *Vernon and Aharoni* 5: 70-84.

Gustafson, T. (1999). *Capitalism Russian-style.* Cambridge: Cambridge University Press.

Harberger, A. (1971). "Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay." *Journal of Economic Literature* 9 (3): 785-797.

Hart, O. and Moore, J. (1990). "Property Rights and the Nature of the Firm." *Journal of Political Economy* 98 (6): 1119-1158.

Hart, O., Shleifer, A., and Vishny, R. (1997). "The Proper Scope of Government: Theory and an Application to Prisons." *Quarterly Journal of Economics* 112: 1127-1161.

Hayek, F. (1944). *The Road to Serfdom.* Chicago: University of Chicago Press.

Hayek, F. (ed.) (1935). *Collectivist Economic Planning: Critical Studies on the Possibilities of Socialism.* London: G. Routledge & Sons.

Hirschman, A. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States.* Cambridge, MA: Harvard University Press.

Kindleberger, C. (1984). *A Financial History of Western Europe.* London: George Allen & Unwin.

Kingston, W. (2000). "A Spectre is Haunting the World—the Spectre of *Global* Capitalism." *Journal of Evolutionary Economics* 10: 83-108.

Little, I. M. D. (1957). *A Critique of Welfare Economics.* Oxford: Clarendon Press.

López-de-Silanes, F., Shleifer, A., and Vishny, R. (1997). "Privatization in the United States." *Rand Journal of Economics* 28 (3, autumn): 447-471.

Lutz, H. (1930). *Public Finance.* New York: D. Appleton and Company, 2nd ed.

Madsen, A. (1916). *The State as Manufacturer and Trader.* London: T.F. Unwin, Ltd.

Martinelli, A. (1981). "The Italian Experience: A Historical Perspective." *Vernon and Aharoni* 6: 85-98.

Mises, L. (1935). "Economic Calculation in the Socialist Commonwealth." Reprinted in Hayek (1935).

Mitchell, Wesley C. (1951). *What Happens during Business Cycles: A Progress Report.* New York: National Bureau of Economic Research.

Moe, T. (1997). "The Positive Theory of Public Bureaucracy." In *Perspectives on Public Choice: A Handbook,* Mueller, Dennis C., ed., Ch. 21. Cambridge and New York: Cambridge University Press.

Mueller, D. (1989). *Public Choice II: A Revised Edition of Public Choice.* Cambridge: Cambridge University Press.

Normanton, E. L. (1981). "Accountability and Audit." *Vernon and Aharoni* 11: 157-169.

Olson, M. (1965). *The Logic of Collective Action.* Cambridge, MA: Harvard University Press.

Olson, M. (1982). *The Rise and Decline of Nations.* New Haven: Yale University Press.

Olson, M. (2000). *Power and Prosperity: Outgrowing Communist and Capitalist Dictatorships.* New York: Basic Books.

Peirce, W. (1981). *Bureaucratic Failure and Public Expenditure.* New York: Academic Press.

Pigou, A. (1962). *The Economics of Welfare* [1920]. London and New York: Macmillan, fourth ed.

Raiffa, H. (1981). "Decision Making in the State-owned Enterprise." *Vernon and Aharoni* 3: 54-62.

Rapaczynski, A. (1996). "The Roles of the State and the Market in Establishing Property Rights." *Journal of Economic Perspectives* 10 (2) Spring: 87-103.

Robinson, J. [1933]. *The Economics of Imperfect Competition.* London: Macmillan, 1965.

Robson, W. (1962). *Nationalized Industry and Public Ownership.* London: George Allen & Unwin Ltd., 2nd ed.

Rosa, Jean-Jacques. (1993). "Nationalization, Privatization, and the Allocation of Financial Property Rights." *Public Choice* 75 (4): 317-337.

Rosegger, G. (1985). "Industrial Policy and State Capitalism: the Austrian Experience." Working Paper. Cleveland: Case Western Reserve University, Economics Department.

Rossi, E. (1955). "Natinalization in Italy." In Einaudi et al., pp. 191-246.

Rowley, C. and Yarrow, G. (1981). "Property Rights, Regulation and Public Enterprise: The Case of the British Steel Industry 1957–1975." *International Review of Law and Economics* 1: 63-96.

Savas, E. (1982). *Privatizing the Public Sector.* Chatham, NJ: Chatham House Publishers, Inc.

Savas, E. (1999). "Competition and Choice in New York City Social Services." Working Paper, Privatization Research Organization, School of Public Affairs, Baruch College, City University of New York.

Schmidt, K. (2000). "The Political Economy of Mass Privatization and the Risk of Expropriation." *European Economic Review* 44 (2): 393-421.

Sheshinski, Eytan and López-Calva, Luis Felipe. (1998). "Privatization and its Benefits: Theory and Evidence." Paper prepared for *Consulting Assistance on Economic Reform (CAER) II* project at the Harvard Institute for International Development.

Shleifer, A. and Treisman, D. (2000). *Without a Map: Political Tactics and Economic Reform in Russia.* Cambridge, MA: The MIT Press.

Shleifer, A. and Vishny, R. (1994). "Politicians and Firms." *Quarterly Journal of Economics* CIX (November): 995-1025.

Siegmund, Uwe (1997). "Are There Nationalization-privatization Cycles?" *Economic Systems* 21 (4): 370-374.

Smith, A. (1776). *An Inquiry Into the Nature and Causes of the Wealth of Nations.* New York: Modern Library edition, 1939.

Taylor, J. B. (1927). "Some Phases of North Dakota's Experiment in Flour Mill Operation." *The Accounting Review* II(2): 129-139. Reprinted in *Readings in Public Finance and Taxation,* Mills, M. and Starr, G., eds., pp. 195-204. New York: The Macmillan Company, 1932.

Tullock, G. (1965). *The Politics of Bureaucracy.* Washington: The Public Affairs Press.

Van Dorn, H. (1926). *Government Owned Corporations.* New York: Alfred A. Knopf.

Van Hise, C. (1912). *Concentration and Control: A Solution of the Trust Problem in the United States.* New York: The Macmillan Company.

Vernon, R. and Aharoni, Y. (eds.) (1981). *State-owned Enterprise in the Western Economies.* New York: St. Martin's Press.

Vickers, J. and Yarrow, G. (1988). *Privatization: An Economic Analysis.* Cambridge, MA: The MIT Press.

Warwick, D. with Meade, M., and Reed, T. (1975). *A Theory of Public Bureaucracy: Politics, Personality, and Organization in the State Department.* Cambridge, MA: Harvard University Press.

Yergin, Daniel and Stanislaw, Joseph (1998), *The Commanding Heights: The Battle Between Government and the Marketplace that is Remaking the Modern World.* New York: Simon & Schuster.

Chapter 13

SOCIAL INSURANCE

Heinz Grossekettler
University of Münster
grossekettler@wiwi.uni-muenster.de

Abstract   This chapter deals with some major problems associated with social insurance systems. After examining the reasons which led to the emergence of these systems, the author attempts to shed some light on their characteristic growth pattern. He subsequently undertakes a comparative study of social insurance systems in OECD member states, discusses some difficulties which have arisen within the European Single Market, gives a general overview of the economic effects of social insurance, and pinpoints problems posed by population aging and various proposals for pension reform.

1.   DEFINITIONS AND HISTORY

1.1.   Social Insurance, Social Security, and the Welfare State

The terms *social insurance, social security* (or *protection*) and *welfare state* are sometimes used to denote different concepts, but they are often employed as synonyms. When they are not used interchangeably, *welfare state* is generally the term with the widest range of meaning: it is a superordinate of *social security,* just as *social security* is a superordinate of *social insurance,* the latter term serving to denote the heart of the welfare state, i.e., the most important system of protection schemes.

When the term *welfare state* was coined in 1941, it was a political slogan with positive connotations: "Archbishop Temple coined the phrase to differentiate wartime Britain from the 'warfare state' of Nazi-Germany. It quickly entered the vocabulary associated with the Beveridge Report (1942)" (Gough, 1998, p. 895). In the Anglo-Saxon literature on the subject the following definition by Briggs (1961, p. 228) has attained wide currency:

> A "Welfare State" is a state in which organised power is deliberately used (through politics and administration) in an effort to modify the play of market forces in at least three directions—first, by guaranteeing individuals and families a minimum income irrespective of the market value of their property; second by narrowing the extent of insecurity by enabling individuals and families to meet certain "social contingencies" (for example, sickness, old age and unemployment) which lead otherwise to individual and family crises; and third by ensuring that all citizens without distinction of status or class are offered the best standards available in relation to a certain agreed range of social services.

This definition outlines the main characteristics of the welfare state, but it can be interpreted in a wider or a narrower sense. In German, by contrast, these different senses are rendered by two distinct terms (Wendt, 1993, pp. 29-32). The term *Wohlfahrtsstaat* denotes a state endowed with social protection and redistribution schemes which are so highly developed that the state assumes a paternalistic character, stifling private initiative and creating an excess burden. The correlative term, *Sozialstaat*, carries more positive connotations and has a narrower range of meaning. The *Sozialstaat* represents a limited form of welfarism which belongs to the foundations of the German state and is anchored in the constitution (*Grundgesetz*,[1] Article 20, §1).

We can bring the concept of the welfare state into sharper focus by enumerating its constituent elements. The following Table 1, which was compiled by Flora and Heidenheimer (1982) and adapted by Gough (1998, p. 895), illustrates a widely accepted enumerative definition to which many authors add a full-employment policy, poor relief (or social assistance) and public housing for the disadvantaged.[2] The table also shows that the welfare state came into being at the turn of the nineteenth into the twentieth century.[3]

The term *social security (system)* first crops up in the American Social Security Act enacted in 1935. Roosevelt assigned social security as a category comprising four basic freedoms: freedom of worship, freedom of speech and expression, freedom from want, and freedom from fear (Weisser, 1956, p. 397). As a rule, the term *social security* denotes the various types of social insurance listed in Table 1, in addition to certain forms of poor relief or social assistance.

When the expressions under discussion are used in different senses, *social insurance* generally denotes the core of the welfare state as well as the core of the social security system, i.e., the elements comprised in the upper part of Table 1. However, the terms used to denote protection schemes differ from

TABLE 1.
Introduction of Welfare State Elements

Year of Introduction of:	Germany	U.K.	Sweden	France	Italy	U.S.A.	Canada
Industrial accidents	1884	1906	1901	1946	1898	1930	1930
Sickness insurance	1883	1911	1910	1930	1943	–	1971
Pensions	1889	1908	1913	1910	1919	1935	1927
Unemployment insurance	1927	1911	1934	1967	1919	1935	1940
Family allowances	1954	1945	1947	1932	1936	–	1944
Health insurance/ service	1880	1948	1962	1945	1945	–	1972
General personal income tax operated	1920*	1918	1903	1960	1923	1913	N/A
uninterruptedly from:	(1873)						
Education:							
Adult illiteracy first < 20%	1850	1880	1880	N/A	N/A	1870	N/A
Secondary school enrolment first > 10%	1925	1923	1937	N/A	N/A	1915	N/A
University enrolment first > 10%	1975	1973	1968	N/A	N/A	1946	N/A

*Introduced in Prussia in 1873.

Source: Gough (1998, p. 895) adapted from Flora and Heidenheimer (1982).

country to country. The same is true of the organisational structures of these institutions, the demarcation of their respective spheres of responsibility and the extent of the risks they cover.

In order to avoid definitional problems, we shall henceforth use *welfare state, social security* and *social insurance* as synonyms. The main emphasis, however, will be on social insurance systems in the strict sense. In Section 2 ("Social Insurance in Modern Times") we shall make some international comparisons. Before that, however, I would like to look briefly at the motives and functions in which social insurance schemes originated. This historical survey may throw interesting light upon the driving forces at work in this domain, and it may also conduce to a better understanding of the emergence and the growth patterns of modern social insurance systems.

1.2.   Explanations for the emergence of social insurance systems: motives and functions

In what now follows, we shall formulate a number of hypotheses about the emergence of the more or less extensive social security systems which have been set up in human societies.

Owing to a lack of sufficient data, and for several other less important reasons, these hypotheses cannot be verified by statistical methods. Nonetheless, as the next subsection will show, the plausibility of our hypotheses can be tested with the aid of research findings published by historians and natural scientists.[4]

Let us assume the existence of a state where considerable importance is attached to the citizens' votes, or where much depends on the government's reputation and the support the government receives from the citizens. Our hypotheses about this kind of state can be divided into two groups, each consisting of at least two elements.

The hypotheses assigned to the first group are based on the assumption that social security systems were set up in order to meet the individual needs of citizens and/or those in power. The elements in this group will henceforth be referred to as "theories oriented towards subjective needs" or (more briefly) as "need-oriented theories."

By contrast, the hypotheses placed in the second group are based on the assumption that economic systems can be made to function more effectively. The postulates in this group will be described as "function-oriented theories."

Need-oriented theories can be used directly as an aid to understanding why private or public initiatives may be launched in order to meet certain needs within a democratic system. Function-oriented theories, by contrast, fail to provide a sufficient explanation for such initiatives. In the view of public-choice theorists, function-oriented theories have to be supplemented by an additional hypothesis, which might be formulated as follows: "In a democratic state there must be people who are capable of convincing the majority of their fellow citizens that it is in the best interests of the nation to make the economic system function more effectively. Figuratively speaking, these people assume the role of a political entrepreneur whose influence might be likened to that of a charitable dictator. However, it would clearly be an error to take it for granted that such political entrepreneurs can be found in every democracy or that people with such moral fiber will invariably succeed in asserting themselves."[5]

In what follows, we shall use the term *moral sentiment motive* to denote the driving force whose existence is postulated in the aforementioned need-oriented hypotheses. In *The Theory of Moral Sentiments* the force in question is attributed to an individual need which Smith calls *sympathy* (Smith,

1976/1759). In Smith's view, our capacity for sympathy generates a spontaneous readiness to help others. Although egoism is the principal motivating force in normal economic life, an impartial observer of human behavior cannot fail to notice that people are not indifferent to the fate of their fellow beings and are prepared to help others even in cases where they will not receive anything in return.[6]

There is a second need-oriented hypothesis which has an equally long tradition and which is based on the idea that social policy serves to hold society together. If we are to believe the proponents of this hypothesis, social policy has a particularly important role to play whenever a society is split up into classes and is in danger of falling apart. Zwiedineck-Südenhorst (1911, p. 36 ff.) expresses the same idea in slightly different terms when he says that social policy is essentially a policy which is developed by the ruling class, and which is designed to prevent a deepening of the rifts within society, mainly by supporting the disadvantaged. Lorenz von Stein (1855) had been even more explicit when he called for the establishment of a "social monarchy," i.e., a state which would be immune to internal unrest because it would offer advantages to every class. The privileged would not need to worry about the danger of a revolution fueled by social tensions, while the disadvantaged would enjoy freedom from want and would be able to make their way up the social ladder. In what follows, the essence of this hypothesis will therefore be termed the *power-safeguarding motive.*

The first group of function-oriented hypotheses is concerned with economic advantages which may be gained by correcting market failure[7] in insurance markets. The proponents of these hypotheses refer to circumstances which hamper or hinder the development of insurances which citizens desire for egoistic reasons or because of the moral sentiment motive. If private insurances are unable to provide a low-cost solution to such problems, the offer of social insurance is a kind of rational social risk management (Diamond, 1977, 1998; Rolf et al., 1988; Rosner, 2003). These theses are foreshadowed by Zwiedineck-Südenhorst (1911, p. 377 ff.).

Classical market-failure arguments hinge on the idea of an asymmetrical distribution of information. Typical examples are provided by the adverse selection, moral hazard and merit good arguments.

We speak of *adverse selection* when an insurer is unable to obtain sufficient information about the types of risk (good or bad) which a prospective policyholder is likely to incur. In this kind of situation an insurer may find that all his potential clients are people who want to be insured against bad risks. As a result, he is obliged to charge premiums which are deemed unattractive by those who require insurance cover against good risks. In extreme cases no one is prepared to create an insurance market (cf. Rosner, 2003, pp. 40-44). When

this kind of problem arises, the state can spread the risks by setting up a compulsory insurance system, thereby creating an insurance market characterized by what is known as a pooling equilibrium. This equilibrium, however, is not Pareto-optimal since it is based on redistribution in favour of bad risks.

There is a *moral hazard* when, after signing an insurance contract, an insuree is tempted to take greater risks than before, or might even toy with the idea of hoodwinking the insurer. This kind of situation may arise when the insurer is unable to keep tabs on the policyholder and check whether he is abiding by the terms of the contract. The government can remedy such problems if citizens are prepared to modify public law in such a way as to allow the state to wield supervisory control of a kind they would deny to private insurers for fear that companies might abuse their power. A suitable example is afforded by obligations imposed on unemployed people—obligations which only the state has the right to decide.

Insurances constitute a *merit good* when there are grounds for believing that people underestimate certain risks—threats which are unlikely to materialize or which will normally materialize only in the very distant future.[8] In such cases insurance contracts tend to be signed too late, and premiums tend to be underpriced. Such errors of judgment are generally committed by young people, who won't be able to correct these mistakes when they are mature and more experienced. The government may seek to remedy this kind of situation by instituting a compulsory insurance system or by introducing a tax-financed minimum insurance.

An examination of recent specialist literature shows that classical market-failure arguments may be complemented by three non-classical insurance-related arguments:

- The first of these arguments hinges on the idea that uninsurable risks include (a) situations where couples incapable of having children fail to receive their pension from their own children (Sinn, 1998, 2000) the possibility that women will give birth to children who are handicapped or of below average intelligence (Sinn, 2000, p. 20).
- The second argument is that social policy can compensate for disadvantages suffered by people who are denied a good start in life because they come from a deprived working-class background. The drawbacks in question are like obstacles to development, and it is impossible to take out an insurance policy against such handicaps (Rolf et al., 1988, p. 32 f.).
- The third argument is that social insurance can remedy a prisoner's dilemma situation in a society where there is a consensus that assistance should be provided for people who find themselves in straitened circumstances through no fault of their own. It is conceivable that the better-off will take it for granted that other well-heeled people will provide financial assistance for the destitute. Yet if a ballot were to be held, there can

be no doubt that the rich would vote unanimously in favour of measures which would oblige wealthy people to provide financial support for the poor. This kind of situation can be resolved by means of compulsory payments (a kind of insurance for people who are reduced to poverty through no fault of their own). However, compulsory payments may result in a Samaritan's dilemma, for people may be tempted to be improvident if they know they will receive assistance in time of need.

All the aforementioned arguments concern government measures intended to enhance security by seeking new ways of providing insurance in the broadest sense. In what now follows, the first group of function-oriented hypotheses will therefore be lumped together in a category labeled *insurance-promoting function.*

The function-oriented hypotheses in the second group concern forms of market failure not directly related to insurance markets. Thus, for instance, Sinn (1998, p. 20 f.) stresses that since social policy increases people's willingness to take risks, it can be regarded as a productive factor capable (within limits) of accelerating growth. Similarly, Ott (2000, p. 191 ff.) emphasizes that social policy plays an important role in building an optimal stock of human capital.

Function-oriented hypotheses are prefigured in older publications. A good example is provided by the so-called state insurance theory (Rolf et al., 1988, p. 15 f.).

In what follows, the goal in question will be referred to as the *growth-promoting function.* In discussions concerning the insurance-promoting goal, distributive considerations often acquire appreciable significance. The growth-promoting goal, by contrast, is mainly allocative in character.

1.3.   What experience shows: the views of natural scientists and historians

In the preceding section we put forward a number of hypotheses about the emergence of social security systems. Now we shall consider to what extent these hypotheses are compatible with the findings of ethologists and historians.

(1) The following considerations may shed some light on the moral sentiment motive:

- Ethologists point out that the behavior of primates[9] is determined by two related instincts which can be considered as genetically programed (Eibl-Eibesfeldt 1997, p. 482 ff.). The first of these instincts is a peacemaking removal inhibition (i.e., a concept of property). The second instinct is a tendency which originated in a

brood-care system involving the feeding of offspring—a tendency to give something away if the recipient demonstrates fundamental respect for the giver's property. This second tendency corresponds to the moral sentiment motive: If we are to believe Eibl-Eibesfeldt (ibid.), primates sometimes hold out their hands as if they were begging—a gesture which is supposed to fulfill a double function. On the one hand, it activates the moral sentiment motive, while on the other hand it serves to express peaceableness and respect for property. In humans, the removal inhibition instinct and the willingness to share certain things with others coalesce, thereby generating a propensity to exchange goods (trade). This propensity can take on a variety of forms: an exchange of goods based on mutual egotism, an exchange of friendly gestures (e.g., the offering of presents), and mutual support within groups. Thus, for instance, in order to gain the support of people from the lower echelons of society, those who belong to the upper echelons offer their inferiors a share of the spoils of the chase.

- The findings of sociobiologists (Wilson, 1976, p. 120 f. and 551 f.) lend support to these ideas.
- Anthropologists such as Harris (1995, p. 21 ff.) point out that even the emergence of states can be explained by the aforementioned instincts. In almost every primitive society we find "great men" who win other people's allegiance by distributing presents, thereby forming bands out of which power structures evolve.
- Finally, Frank (1992) points out that in certain circumstances "moral behavior" may assume the character of an instrument for solving economic problems which cannot be resolved by out-and-out egoists. In his view, it is therefore possible to gain *individual* selection advantages by yielding to the promptings of the moral sentiment. These ideas constitute a theory of selection which—unlike the theory of selection advantages for *groups* acting in accordance with ethical principles—is perfectly compatible with the findings of biological research. Ridley (1996) expresses a similar idea when he says that it often pays to be "good" in an egoistic sense.

All in all, therefore, the research findings presented by natural scientists suggest a thesis which—from the economic viewpoint—is quite remarkable. The essence of this thesis is the idea that two diametrically opposed phenomena have common roots in the genetic program of the human race.

The common roots of these phenomena may be classified as follows: (1) a strong desire to reserve acquisitions permanently for one's own use

and prevent others from gaining access to them; (2) the ability to recognize other people's property in order to obtain a "peace dividend," and the capacity to organize useful exchanges on this basis; (3) willingness to support other members of a group;[10] (4) the urge to win the esteem of one's fellow creatures by doing philanthropic work; and finally—if we take account of religious beliefs—(5) the attempt to buy salvation in the hereafter by performing acts of charity in this world.

A number of historians have concerned themselves with the development of sociopolitical institutions, and their findings lend support to the thesis that the moral sentiment motive is one of the constants of human behavior.

Reports on early social measures and institutions can be found in Frerich and Frey (1996a, p. 1 ff.), Roscher (1894, p. 67 ff.) and Zwiedineck-Südenhorst (1911, p. 67 ff.). Most of the information furnished by these authors concerns the ancient Orient, Israel, ancient Greece, ancient Rome, early Christian societies, and the Germanic peoples. However, it is impossible to determine whether or to what extent the measures in question are motivated exclusively by the moral sentiment.

In recent years various attempts have been made to test the so-called "ethical voter" hypothesis (cf. Volkert, 1999, p. 105). These studies have shown that social welfare programs are not only popular with those who benefit by them; they go down well with voters in general. Compelling evidence supporting the moral sentiment hypothesis is also provided by the willingness to help others that can be observed whenever a disaster occurs; and the same goes for the selflessness evinced by soldiers rescuing wounded comrades or by firemen and police officers risking their lives in burning buildings.

Despite this evidence, economists are generally reluctant to invoke the moral sentiment motive. Their reluctance is mainly due to methodical reasons. To be more precise, they are wont to apply the principle known as Ockham's razor. Mindful of the rule that "plurality should not be assumed without necessity," they are anxious to ensure that the concepts and motives they invoke should be reduced to a minimum, and they feel they are on safer ground saying that human behavior can be explained by amour propre.[11] This methodical reserve is basically healthy, but—as we shall demonstrate—it must be abandoned whenever it is necessary to invoke the moral sentiment motive in order to achieve an adequate understanding of certain developments.

(2) In order to furnish empirical proof of the importance of the *power-safeguarding motive,* we have already cited research findings reported

by natural scientists. Now we shall bring our attention to bear on his-
torical sources:

- Let us begin by recalling the thesis put forward by natural sci-
  entists, namely the idea that higher ranking persons try to create
  power structures by distributing presents to lower ranking persons
  whom they wish to make dependent on themselves.
- Social policy has often been used in order to consolidate power
  structures. Historical evidence of this was adduced in early so-
  ciopolitical publications such as Zwiedineck-Südenhorst (1911,
  p. 67 ff.) and even Roscher (1894, p. 67 ff.). Striking examples are
  provided by the social legislation of Hammurabi (1728–1686 BC)
  and the practice of buying proletarians' votes in ancient Greece
  and ancient Rome.
- Having pursued a successful policy of conquest, Hammurabi built
  a multiethnic state torn by severe internal tensions. In order to al-
  leviate these tensions, he adopted a policy designed to bring peace
  to his realm. His main policy instrument was a series of social
  measures intended to convince all his subjects that it was in their
  best interest to remain under his protection (Klengel, 1999, p. 62).
- The word *proletarian* is derived from the Latin noun *proles,* which
  denoted citizens who were not landowners, but who had the right
  to vote because they were the descendants of full citizens. In the
  ancient democracies such people often benefited from social mea-
  sures. The best known example of this is the policy pursued in
  ancient Rome in the wake of the riots sparked off by the Grac-
  chi's attempts at land reform (133 and 121 BC). The Roman ruling
  class sought to quiet the proletariat by spectacles and by doles of
  bread—a social policy which Juvenal referred to contemptuously
  in the trenchant phrase *panem et circenses.*
- The best-known example of the power-safeguarding motive is no
  doubt Bismarck's introduction of social insurance in Germany. As
  we have already pointed out, Lorenz von Stein (1855) had devel-
  oped the idea of a "social monarchy," i.e., a state which would
  offer something to every social class. It has been proved that the
  idea in question found a staunch supporter in Hermann Wagener,
  who provided Bismarck with some of his most important sociopo-
  litical ideas (Saile, 1958). According to a historical survey pub-
  lished by the Deutscher Bundestag (1996, p. 183 f.), Bismarck
  wanted to give the vast propertyless class that conservative cast of
  mind which is produced by the experience of retirement. Histori-
  ans agree that German social legislation must be viewed in con-

nection with the so-called Socialist Laws, which were designed to combat social democratic organizations:

"At any rate, contemporaries perceived a close relationship between the Socialist Law and the laws on workers' insurance. This explains why, in the 1880s, social democrats roundly rejected social insurance as the expression of a 'carrot and stick' policy." (Ritter, 1997, p. 692 f.)

- Nonetheless, Lampert (1998, p. 42 ff.), Craig (1999, p. 176 ff.) and other historians stress that it would be a mistake to attribute social insurance laws exclusively to the power-safeguarding motive. According to these authors, it is preferable to assume a combination of the moral sentiment motive and the power-safeguarding motive.
- English history also affords evidence of the relevance of the power-safeguarding motive. This becomes clear when we ask why social insurance benefits were introduced so much later in Great Britain than in Germany, and why the British system was modeled on the German one although Great Britain's economy was then more advanced than Germany's. The answer given by historians is that Great Britain—unlike Germany—did not have a labour movement that was considered a threat to the entire social system (Ritter, 1983, p. 77). It is also significant that when the vulnerability of the British Empire was revealed by the second Boer War (1899–1902), England's political parties called unanimously for the modernization of the country. It remains to add that measures designed to improve the health, education and social security of the working class were invariably viewed as an essential precondition for a more effective defense of the Empire (Ritter, 1983, p. 82 f.).

(3) Although we have examined historical sources as well as literature on natural sciences, we have been unable to find any convincing examples of the insurance- or growth-promoting functions.

By applying the methods of sociobiological or comparative behavioral research, it ought to be possible to demonstrate that whenever there is scope for stronger growth or improvement in insurance markets, that potential is fully exploited in a manner which is reminiscent of the innovative and imitative processes observable in commodity markets. This cannot be taken for granted, for the improvements we have in mind cannot be effected by isolated individuals; they require a collective decision-making process that presupposes insights into significant relationships. Philosophers, political scientists, sociologists—and even some economists—have occasionally asserted that such improvements

are indeed effected, but—to the author's knowledge—the methods of natural science have not yet been applied to the study of the ameliorations in question.[12]

The historical sources examined by the author do not contain any information suggesting that insurance- and/or growth-promoting functions as such have been responsible for sociopolitical initiatives. Works devoted to the history of insurance contain indications that guilds once fulfilled functions similar to those now assumed by social insurance (Schewe, 2000, pp. 133-136, etc.), and some people have complained that opportunities to improve such institutions were missed. Zwiedineck-Südenhorst (1911, p. 386 ff.) discusses health-insurance companies and social welfare funds which were set up on a voluntary basis and had a local catchment area; he says that such institutions often failed because they were mismanaged and too small to benefit from the law of large numbers.[13] These two fatal flaws can probably be explained by the fact that tendencies towards adverse selection and moral hazard could only be checked at the local level because all the persons concerned knew each other and social constraints assumed a conspicuous role. Thus, while there is clear evidence of the type of market failure we have just described, there are no grounds for believing that market failure alone provided an incentive to set up a social insurance system.

All in all, there is a sufficient volume of empirical evidence to indicate that sociopolitical initiatives have been determined not only by the moral sentiment motive, but also by the power-safeguarding motive. Sometimes the two motives have acted in unison, while on other occasions one of the motives has acted in isolation.

There have been many opportunities to promote growth and evolve better insurance systems. Such opportunities may be seized upon by economists looking for plausible explanations for various sociopolitical phenomena, and they may also be used to bolster up the arguments deployed by politicians who are anxious to defend the social policies they have framed. Nonetheless, it must be stressed that the sociopolitical intiatives we have just referred to have never been triggered by growth opportunities or by a juncture of circumstances favouring improvements in existing insurance schemes.

In order to explain the historical development of social security systems, and in order to elucidate the political forces that have shaped them, we shall therefore have to concentrate on the moral sentiment and power-safeguarding motives. In the next section we shall attempt to demonstrate that the interaction of these two motives inevitably produces a specific growth pattern which is amenable to empirical observation.

2.   SOCIAL INSURANCE IN MODERN TIMES

2.1.   The Growth Pattern and the Driving Forces
behind it: The Public Choice View

Although there are marked differences between the ways in which social
security systems have been instituted in modern states, we can say that the
systems in question have gone through a three-phase development in all the
countries under discussion. There is an *early phase* during which the system
is set up and elaborated in such a way as to make provision for the various
types of risk that are insured. Then comes an *expansion phase* during which
the social budget consumes a growing amount of GDP as more generous social
benefits are provided to a broader section of the population. Finally, there is a
stagnation phase characterized by worsening financing problems.[14]

As can be seen from Table 1 (p. 325), there are major industrial states in
which the early phase did not end until after World War II. Our table, however,
fails to reveal any crucial changes. In 1957, for instance, the German pension
insurance system was rendered more dynamic when a link was established
between pension insurance and wage development; in addition, the funding
system was replaced by the pay-as-you-go system, and in 1994 a "fifth pillar"
was added to the welfare system built around accident, health, pension and
unemployment insurance. This fifth pillar is nursing-care insurance *(Pflegev-
ersicherung).*

Since space precludes detailed treatment of the specific features of all the
major state social insurance systems, we shall confine our attention to develop-
ments in Germany between 1960 and 2000. The emphasis will be firmly placed
on points of resemblance between the German system and systems erected in
other countries.

Germany offers an excellent example for several purely statistical reasons:

- Given the advanced age of the German social insurance system, Germany
  can be justly described as a "mature" welfare state.
- As our comparisons will demonstrate, the level of social welfare benefits
  is, on the whole, much higher in Germany than in many other states.[15]
- Comparisons with other welfare states show that Germany is by no means
  exceptional. Since the relationship between its per capita social spending
  and its per capita GDP is perfectly normal (Alber, 1999, pp. 44-49), it
  can safely be stated that Germany occupies an intermediate position be-
  tween the high-spending, state-centred Scandinavian welfare states and
  their low-spending Anglo-Saxon counterparts, which rely heavily on pri-
  vate initiative (Alber, 1999, p. 62).
- We have statistical series for Germany, and these series go back farther
  into the past than the series which are available for most other countries.

From a purely technical viewpoint, this is particularly important for the analysis of development patterns.

- If we look closely at welfare benefit data for the period under review, we can observe an expansion phase and a stagnation phase. This is equally true of Germany and of many other advanced industrial countries.[16]
- German reunification (1990) was a historic event which Peacock and Wiseman (1961) would have regarded as a "social disturbance."

On the basis of these data, we shall adduce evidence of correlations which probably have universal validity. We shall also discuss parallels and differences between our own theory and the hypotheses put forward by Peacock and Wiseman (1961) in order to account for increases in the public sector's share of GNP. Recent developments in other countries will be treated in the next main section.

Apart from the two oil crises, Germany and the other big industrial states were not exposed to any major exogenous shocks between 1960 and 1989. As a result, economic development was mainly determined by endogenous forces.

In what now follows, we proceed on the assumption that the development of social spending is determined by the interaction of exogenous shocks and permanent endogenous forces. This basic hypothesis suggests that it may be worthwhile to scrutinize phases in which development is determined primarily by endogenous factors. A searching study of such phases should reveal the forces which are constantly at work in major industrial states.

The period between 1960 and 1989 may be divided schematically into two sub-periods corresponding to two distinct phases of development. These phases are linked by a transitional period (1977–1980) whose development patterns reveal, among other things, the impact of the oil crises.

As far as social policy is concerned, the first phase (1961–1976)[17] is characterized by a *rapid increase* in quotas. The second phase (1981–1989),[18] by contrast, is characterized by *stagnation and a partial reduction in quotas.* These two phases are represented in Figure 1.

The main features of the *first phase* are as follows:

- The growth rates of social spending and GDP are very high. The median growth rate of social spending is over 11%, while the corresponding GDP figure is almost 9%.
- The growth rate of social spending is higher than the corresponding GDP figure. This leads to an increase in the social expenditure ratio.
- At a low level of significance, there is a positive correlation between the two growth rates under consideration. The correlation coefficient is somewhere in the region of 0.31.
- The unemployment rate is relatively low (almost always below 4%).

1a) Growth Patterns for Social Spending (**━**) and GDP (**━**) and the Evolution of the
     Unemployment Rate (**━**)*
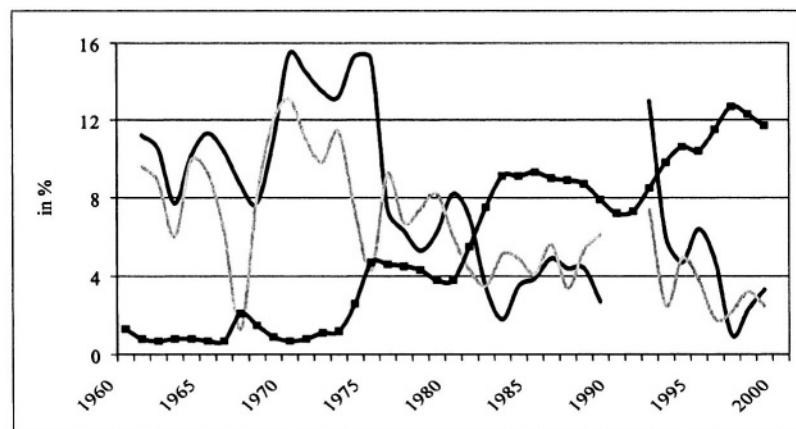
1b) The Evolution of the Social Expenditure Ratio (**━**) and the Slope of the Line of Support (**━**)
     resulting from the Displacement
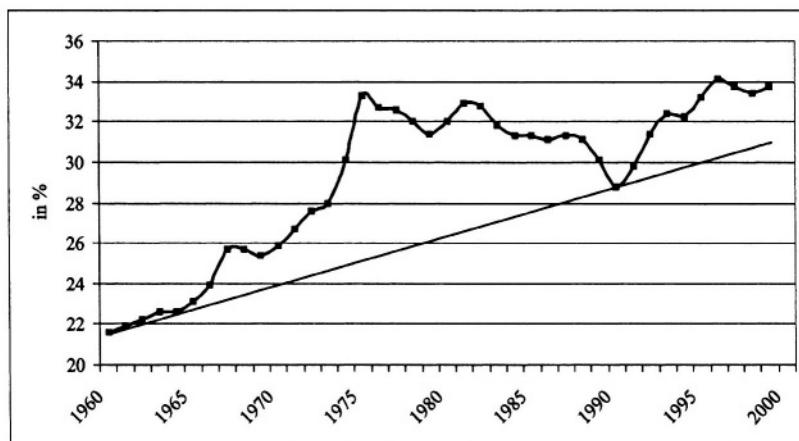
FIGURE 1. Growth Characteristics of Social Spending in Germany.
* The reference numbers for these patterns can be found in Table 2 (p. 16).

By contrast, the principal characteristics of the *second phase* are as follows:

- The growth rates of social spending and GDP are puny. The median
  growth rate of social spending is a mere 4%, while the corresponding
  GDP figure is slightly less than 5%.

TABLE 2.
Reference Numbers for the Evolution of the German Social Budget[a]

	Social Budget				Gross Domestic Product[c]		
Year	DM Bn	Change in %	Social Expenditure Ratio[b,c]	Per capita in DM	DM Bn	Change in %	Unemployment Rate in %
1960	65.3		21.6	1177	302.7		1.3
1961	72.6	11.2	21.9	1292	331.7	9.6	0.8
1962	80.2	10.5	22.2	1411	360.8	8.8	0.7
1963	86.4	7.7	22.6	1506	382.4	6.0	0.8
1964	95.1	10.1	22.6	1641	420.2	9.9	0.8
1965	105.9	11.3	23.1	1807	459.2	9.3	0.7
1966	116.9	10.4	23.9	1977	488.2	6.3	0.7
1967	127.1	8.7	25.7	2143	494.4	1.3	2.1
1968	136.8	7.7	25.7	2299	533.3	7.9	1.5
1969	151.6	10.8	25.4	2523	597.0	11.9	0.9
1970	174.8	15.4	25.9	2883	675.1	13.1	0.7
1971	200.2	14.5	26.7	3266	749.5	11.0	0.8
1972	227.3	13.5	27.6	3685	823.1	9.8	1.1
1973	257.2	13.2	28.0	4149	917.3	11.4	1.2
1974	296.5	15.3	30.1	4777	983.8	7.3	2.6
1975	341.4	15.2	33.3	5521	1026.0	4.3	4.7
1976	366.5	7.4	32.7	5956	1120.3	9.2	4.6
1977	389.6	6.3	32.6	6345	1195.0	6.7	4.5
1978	410.3	5.3	32.0	6690	1283.4	7.4	4.3
1979	435.5	6.2	31.4	7098	1388.4	8.2	3.8
1980	471.3	8.2	32.0	7654	1471.5	6.0	3.8
1981	504.2	7.0	32.9	8174	1534.8	4.3	5.5
1982	521.4	3.4	32.8	8459	1588.1	3.5	7.5
1983	530.5	1.8	31.8	8637	1668.5	5.1	9.1
1984	548.9	3.5	31.3	8972	1750.9	4.9	9.1
1985	570.1	3.9	31.3	9342	1823.2	4.1	9.3
1986	597.8	4.9	31.1	9790	1925.3	5.6	9.0
1987	623.9	4.4	31.3	10215	1990.5	3.4	8.9
1988	651.1	4.4	31.1	10597	2096.0	5.3	8.7
1989	668.9	2.7	30.1	10778	2224.4	6.1	7.9

- As a rule, the growth rate of social spending is lower than the corresponding GDP figure. This leads to a decline in the social expenditure ratio.
- At a low level of significance, there is a negative correlation between the two growth rates under discussion. The correlation coefficient is approximately −0.24.
- The unemployment rate is relatively high. It is generally over 8%.

TABLE 2.
Continued

Year	Social Budget				Gross Domestic Product[c]		Unemploy-ment Rate in %
	DM Bn	Change in %	Social Expenditure Ratio[b,c]	Per capita in DM	DM Bn	Change in %	
1990	725.9	–	28.8	–	2522.8	–	7.2
1991	875.6	–	29.8	10947	2938.0	–	7.3
1992	989.4	13.0	31.4	12276	3155.2	7.4	8.5
1993	1048.6	6.0	32.4	12917	3235.4	2.5	9.8
1994	1097.4	4.7	32.3	13479	3394.4	4.9	10.6
1995	1167.9	6.4	33.2	14302	3523.0	3.8	10.4
1996	1224.2	4.8	34.1	14949	3586.5	1.8	11.5
1997p	1236.8	1.0	33.7	15073	3666.5	2.2	12.7
1998p	1265.1	2.3	33.4	15423	3784.4	3.2	12.3
1999p	1306.6	3.3	33.7	15917	3877.2	2.5	11.7

[a] Up to 1989 the figures refer exclusively to West Germany. From the second half of 1990 onwards, however, the new Länder are included. As a result, post-1990 figures can only be partly compared with those for previous years.

[b] Social security benefits as percent of gross domestic product.

[c] From 1991 onwards GDP has been calculated in accordance with the new European National Accounting System (1995). As a result, absolute figures representing social expenditure ratios can only be partly compared with those for previous years.

Sources: Bundesministerium für Arbeit und Sozialordnung, Sozialbudget 1999, Table I-1; Statistisches Taschenbuch 2000, Arbeits- und Sozialstatistik, Table 2.10. Available on the Internet http://www.bmgs.bund.de/downloads/sozialbudget1999.pdf and http://www.bmgs.bund.de/downloads/stat2000.zip

As we shall see, this schematic presentation is—at least outwardly—reminiscent of the way in which Peacock and Wiseman (1961)[19] present the development of the public sector's share of GDP in Great Britain. We shall now endeavour to explain the development in question.

We can pursue two promising lines of inquiry. First, we might have a look at hypotheses concerning various aspects of social policy, including the ways in which policy measures have been shaped by human nature. Second, we might consider a factor which is undoubtedly exogenous in a narrow sociopolitical sense although it is not exogenous in the macroeconomic sense. The factor in question is a growth cycle which might be considered as the continuation of the Kondratieff cycle.

In the phase we wish to examine, the growth cycle consists of a high-growth phase and a low-growth phase. In principle, the growth cycle and social policy are interdependent, but the influence of the growth cycle on social policy is much stronger than the influence that social policy exerts on the growth cycle.[20]

Our point of departure is the public choice hypothesis—an assumption which, as we have already demonstrated, is supported by a considerable volume of empirical evidence. The basic idea is that social measures introduced by democratically elected governments are constantly influenced by two major factors: (1) the moral sentiment motive and (2) the power-safeguarding motive. The moral sentiment motive plays an important part in determining the attitudes adopted by voters. The power-safeguarding motive, by contrast, is the key to understanding the behavioural patterns of policymakers. For the purposes of this study, voters will be considered as demanders and policymakers as suppliers.

Election results are not only influenced by people who benefit directly from welfare measures. A crucial role is also played by electors who set great store by moral principles. This is particularly the case when the country is chalking up a healthy growth rate and the prospects for further economic growth are looking rosy.

In democratic countries those who hold the reins of government are in constant danger of being voted out of power, and in order to avert this danger they repeatedly resort to social measures. Two points need to be made here:

- Politicians who view public issues in a short-term perspective can gain extra votes by launching a raft of social measures demanded by electors.
- It might seem reasonable to exercise restraint in order to create better conditions for growth, but since such restraint only pays dividends in the long term, politicians who are loath to offer generous welfare benefits tend to be penalized by the electorate.[21]

Conclusion 1:

In democratic countries there is a constant demand for social measures, and it is in the policymakers' best interest to satisfy this demand, which tends to rise when the economy is in good shape.

We shall now consider a second phenomenon whose explanatory potential should not be underestimated. The phenomenon in question can be observed again and again throughout the period under review. It is determined by the special nature of many sociopolitical problems and resultant behavioral tendencies which manifest themselves in the political arena. To be more precise, politicians tend to regard many social measures as credit-financed durable capital goods: When welfare measures are brought in, the resultant political benefits[22] are reaped before the end of the parliamentary term during which the legislative bills in question are passed. By contrast, the political damage[23] caused by the new legislation only becomes apparent at a much later date. Indeed, in some cases the consequences of ill-conceived legislation may be suffered by people who were not yet entitled to vote when the laws under discussion were passed.

A further complicating factor is that "rational ignorance" determines voting behavior in democratic countries. This "ignorance" explains the attitudes normally adopted by the government and the opposition when they are confronted with long-term problems like those we have just described. Two points need to be made here:

- As long as the costs entailed by social measures do not appear unduly high,[24] the government and the opposition will vie to offer new welfare benefits.
- However, when the cost of the social safety net begins to spiral out of control,[25] painful reforms have to be carried out.

When sweeping changes are inevitable, politicians are confronted with a particularly thorny problem. If they take the bull by the horns, they may manage to drive through all the necessary reforms, but they may also be voted out of office before they can reap the benefits of their reform efforts.

Regardless of whether they hold the reins of government, politicians tend to procrastinate if they believe the implementation of sweeping reforms might jeopardize their chances of being re-elected. This tendency to stave off change is accentuated by what might be called an "ideological lag." To put it simply, the *Zeitgeist* is slow to change. Most electors are loath to adapt and give up ingrained habits of mind. As a result, politicians only change their behavior when their electors have become aware that the social security system is in urgent need of reform.[26]

Finally, mention must be made of three additional factors which undoubtedly hamper reform: (1) the vested interests of welfare recipients, (2) the vested interests of the officials responsible for administering the social security system, and (3) the fact that taxpayers gradually become accustomed to defraying the cost of a social safety net. The influence exerted by these three impediments might be likened to a ratchet effect.

Conclusion 2:

There are four reasons why social policies are often seriously flawed:

- Politicians deliberately avoid dispelling certain illusions harbored by their electors.
- When the economy is surging ahead, they offer all kinds of electoral gifts in order to outdo their rivals.
- When the economy is in the doldrums, they tend to postpone necessary reform projects.
- Many people are anxious to defend their vested interests and will avoid doing anything that might jeopardize their careers.[27]

The foregoing explanations may be summed up in the form of a hypothesis concerning transitional probabilities viewed in a historical perspective. Our basic assumption may be formulated as follows:

In democratic states governments tend to take social measures whenever certain economic boundary conditions are satisfied.

- When the economy is in good shape, the likelihood is that the social security system will be refined and consolidated. As time progresses, the number of welfare measures brought in will tend to increase.
- When the economy is in poor shape, the social security system begins to run out of money, and cost cuts have to be made. Nonetheless, remedial action tends to be hesitant and insufficient. True, benefits are sometimes restricted, but the cutbacks in question are never so massive as to cancel out all the welfare spending increases that ultimately put the entire social security system at risk. As a result, there is always a displacement effect (a break in the development of public expenditure), and when the relevant data are represented diagrammatically we can see a hump above a line of support, and an upward trend in the public sector's share of GDP (Figure 1).

If we assume that periods of economic expansion will alternate with recessions, asymmetrical transitional probabilities will result in a gradual upward trend. When the economy is expanding, welfare benefits will tend to be more generous, but when the economy is depressed, the social welfare system will cease to develop, or some parts of it will be dismantled in piecemeal fashion.

Since the dawn of the industrial revolution we have witnessed successive periods of economic growth and stagnation, and the resultant patterns look like Kondratieff cycles. The period under review is by no means an exception. Indeed, there is firm evidence that during this period the phenomenon in question occurred not only in Germany, but also in many other advanced industrial countries (Snower, 2000).

At this point, mention must be made of two other phenomena which we have only hinted at so far, and which throw valuable light on the cyclical developments under discussion:

- Social measures promote demand-side growth as long as the unemployment rate is low and investors are optimistic and find the tax burden tolerable. However, there is no escaping the fact that welfare measures mean more economic stagnation when unemployment is high, economic prospects are bleak, and high taxes are regarded as an obstacle to investment (Snower, 2000, pp. 39-44).
- Social insurance is rather like weather insurance. To be more precise, it does not offer genuine insurance cover against independent risks. If

the economy is depressed or demographic trends are detrimental to the
social safety net, so many claims are filed at the same time that the system
begins to run out of money. This is why Snower (2000, pp. 45-48) speaks
of a "quicksand effect." When excessively heavy demands are made on
the welfare state, the system's resources are rapidly overstretched.

These considerations enable us to achieve a better understanding of the de-
velopment pattern represented in Figure 1. Three points merit some comment
here:

- When the GDP growth rate is high and the unemployment rate is low,
  the government can launch a whole raft of social measures which further
  boost economic growth. This explains why social spending grows faster
  than GDP, there is an increase in the social expenditure ratio, and we
  can observe a positive correlation between the growth rates of GDP and
  social spending.
- When the GDP growth rate is low and the unemployment rate is high,
  the welfare system's resources are overstretched, social measures are de-
  layed, and in the long run the welfare state may even be pruned back.
  This explains why social spending now grows more slowly than GDP,
  there is a gradual decline in the social expenditure ratio, and we can ob-
  serve a negative correlation between the growth rates of GDP and social
  spending.
- Owing to the displacement effect, however, the social expenditure ratio
  continues to rise throughout the two periods we have just described. This
  increase is reflected by the quasi trend shown by the line of support.

Now if we are to believe Peacock and Wiseman (1961), the public sector's
share of GDP is only likely to rise when a social crisis erupts. Yet the explana-
tion we have just put forward suggests that in the long run the public sector's
share of GDP will continue to rise even in countries which were not caught up
in the turmoil of two world wars, and which may well have profited from the
conflicts by supplying goods to the belligerents. We contend, moreover, that
the social policies pursued in democratic countries lead to an increase in the
social expenditure ratio, that this increase generally triggers a rise in the pub-
lic sector's share of GDP, and that such phenomena are particularly likely to
occur when the economy is surging ahead. All these contentions are in direct
contradiction to the hypothesis set up by Peacock/Wiseman.

Nonetheless, there is one point on which we are in complete agreement with
Peacock/Wiseman. Owing to the interaction of the moral sentiment and power-
safeguarding motives, the social expenditure ratio is likely to rise whenever
a social crisis erupts. This kind of crisis may be triggered off by a war, a
natural disaster or—in the case of Germany—the need to rebuild an entire state

in the wake of reunification. Since 1990 huge sums have been disbursed in order to provide benefits for disabled people and disaster victims, and the west German social insurance system has been extended to include the inhabitants of the former GDR. The resultant increase in Germany's social expenditure ratio can be explained in a manner which is perfectly compatible with the Peacock/Wiseman hypothesis.

There are therefore two types of situations where we can expect a sharp increase in welfare benefits:

- The first kind of situation is one where the economy is in good shape. Encouraged by strong and steady GDP growth, politicians will vie to offer gifts to the electorate.
- The second kind of situation arises when a government has to meet social challenges which have nothing whatever to do with GDP growth. Good examples of such challenges are provided by events such as wars and German reunification. A government confronted with such momentous events will attempt to resolve a crisis by launching a whole raft of social measures.

In Figure 1b the increase in the social expenditure ratio between 1960 and 1976 is due to the first reason, while the rise in the public sector's share of GDP between 1991 and 2000 is mainly due to the second reason.

If Germany had not been reunified, the economic situation in the early 1990s would have been entirely different, for the rise in the social expenditure ratio $(s = S/Y)$ would have been much lower. We would merely have witnessed a more or less normal increase occasioned by three factors: (1) a looming Europe-wide recession, (2) an increase in spending on unemployment benefits, and (3) a decline in the GDP growth rate $(\dot{s} = \dot{S} - \dot{Y})$. German reunification triggered an economic boom which initially prevented a rise in the social expenditure ratio. Later on, however, the social insurance system had to absorb rising costs occasioned by structural unemployment in the former GDR, and it also had to cope with the enormous problems posed by a flood of new claims when the entire population of the defunct East German state had to be integrated in record time. As a result of this integration process, the social expenditure ratio jumped from about 29% in 1990 to about 34% in 1996. It is only after 1996 that we can observe a return to the trends that were characteristic of the 1980s (stagnation and pruning).

This explanation helps us to understand the developments that took place in Germany between 1960 and 2000. However, it would be an error to conclude that an increase in the social expenditure ratio is inevitable.

Two points need to be made here:

- First, economic forecasters predict a gradual rise in the social expenditure ratio, but their prognosis only applies to economies in which wel-

fare spending is not curbed by institutional mechanisms. So far, such safety devices have never been put in place in any economic system, yet it ought to be possible to rein in profligate social spending by making a few changes to constitutional law in order to set upper limits for social expenditure ratios. Indeed, such measures have already been envisaged in discussions about rates of taxation, the public sector's share of GNP, and the budget deficit's ratio to GDP (Folkers, 1983; Wildavsky, 1980). It must, however, be added that unilateral engagements undertaken in this domain have not been entirely successful. (Wildavsky and Caiden, 2001, pp. 125-151, 291 f.).

- Second, it must be borne in mind that further increases in social expenditure ratios will inevitably be accompanied by a sharp rise in opportunity costs when social spending has already attained a high level. People who advocate further increases in the social expenditure ratio are therefore likely to face stiff opposition from politicians, and the strength of this resistance will probably be proportional to the existing level of social-security payments. In other words, there is a negative correlation between increases in the social expenditure ratio and the size of the ratio, and this negative relation acts as a brake on further spending increases.

2.2. A Comparative Study of Social Insurance Systems in OECD Member States: The Views of Political Scientists

The above analysis was based on the public choice approach. Following typically economic methods, it assumed that the development of the social expenditure ratio is determined not only by policymakers as policy suppliers, but also by ordinary citizens as voters and policy demanders. The empirical basis of the analysis was restricted to Germany.

In what follows we look at a certain number of empirical studies which are based on international comparisons, and which shed some light on recent developments in social security systems. Most of these reports and analyses have been published by sociologists and political scientists. These authors, however, evince little interest in the individual motives of policy suppliers and policy demanders. Instead, they adopt a functionalist approach or set up typologies which serve at least two different functions: (1) they highlight special features of social security systems that tend to occur in clusters, or (2) they exploit the explanatory potential of the roles assumed by specific social groups or institutional configurations. Thus, for instance, special attention is given to the power wielded by labor unions, the influence exerted by the political parties which have held the reins of government at certain periods, and an index

used to measure the clout enjoyed by various political protagonists opposed to change (Siegel, 2000, pp. 14-82; Wagschal, 2000, pp. 52-61).

Unlike the foregoing public choice analysis, these empirical studies do not display that methodological individualism for which professional economists have a marked predilection. Nonetheless, they do bring out the characteristics of modern social security systems, and they spotlight similarities and dissimilarities between organizational structures in various countries.

There is a well-known typology that distinguishes between two types of state: (1) states which tend to bank on self-help (the Bismarck or insurance model), and states which prefer to pin their hopes on state aid (the Beveridge or welfare model). This classification, which is made with the aid of six variables, is presented in Table 3.

TABLE 3.
The Bismarck and Beveridge Models of the Welfare State

	Bismarck Model	Beveridge Model
(1) Beneficiaries	Employees with insurance cover	The entire population
(2) Financing by	Earnings-related contributions	The state budget (taxes)
(3) Cash benefits	Assessed on the basis of lost earnings	Uniform flat-rate benefits
(4) Non-cash benefits	Non-cash benefits from an insurance company or reimbursement of expenses	Provision by a national health service
(5) Administration	Private, often divided equally between employers and employees	Public
(6) Transfer intensity	high	low

Source: closely based on Schmid (1996, p. 59).

As can be seen from the terms employed in the table, this typology reflects the conditions which existed in Germany and Great Britain when their social insurance systems were first set up. The Beveridge model is mainly based on the welfare and provision principles, while the Bismarck model is constructed on the insurance principle. This becomes apparent if we consider some of the constructional principles which underlie social security systems, and which are often mentioned in the specialist literature on the subject under discussion. The primary elements in question are the welfare principle (a guaranteed minimum income for the destitute), the provision principle (a guaranteed minimum income without a means test) and the insurance principle (assessment on the basis of lost earnings).

In the initial period, the main difference between the two countries under discussion was that German and British policymakers were pursuing different goals.[28] The German government was anxious to curb the revolutionary potential of skilled industrial workers. As a result, a pivotal role was assumed by the power-safeguarding motive, and the most obvious solution was to set up an insurance system restricted to workers with revolutionary leanings. In Great Britain, by contrast, the potential for revolutionary action was comparatively limited, and the authorities were mainly interested in social welfare provision for the destitute. Consequently, a predominant role was played by the moral sentiment motive, and the government opted for flat-rate benefits financed by the state budget.

Since their inception, however, the two systems under discussion have evolved considerably, and disparities have been reduced. Systems which originally resembled the Bismarck model now comprise a guaranteed minimum income, while systems that had affinities with the Beveridge model now provide protection against lost earnings (cf. Flora and Alber, 1982, p. 53). Thus, if we attempt to apply the Bismarck versus Beveridge typology to present-day economic realities, it soon becomes apparent that although existing national welfare systems can be classified as Bismarck or Beveridge schemes, such classificatory labels are mere approximations which do not enable us to understand how the systems have evolved or how they have managed to cope with various politico-economic problems.

There is another well-known typology which is based on the assumption that the development of welfare systems is largely dependent on party political factors. According to this hypothesis, we can achieve a better understanding of the way a welfare scheme has evolved if we know something about the political party that enjoyed the longest period of power after 1945 and was therefore able to leave its mark on the system when it was in a decisive stage of development.

The scholars who have proposed this hypothesis stress the following points:

- Regardless of whether they originally had affinities with the Bismarck model or the Beveridge model, all the European welfare systems assumed their current forms after the Second World War.
- Special attention must be given to factors such as a party's ideology and the extent to which workers were split up into many different associations.

The typology based on the party political hypothesis was set up by Esping-Andersen and Korpi (1984), who distinguish three kinds of welfare states: liberal, conservative and Social Democratic states.[29] We can combine this typology with a classification of the typical characteristics of various social secu-

rity systems. The results are shown in Table 4, which we have borrowed from Schmidt (1988, p. 162).

TABLE 4.
Types of Welfare States

Indicator	Type of Welfare State		
	liberal	conservative	Social Democratic
(1) Protection against market forces and loss of earnings	weak	medium	strong
(2) Social rights or poor relief	poor relief	social rights	social rights
(3) Share of private contributions to social expenditure	large	medium-sized	low
(4) Size of the individual's contribution towards expenses	medium-sized	large	medium-sized
(5) Different social security systems for different professional groups	no	yes	no
(6) Redistributive capacity	medium	medium	high
(7) Full employment guarantee	no	only when the economy is in good shape	yes

Source: Schmidt (1988, p. 162).

Broadly speaking, Sweden and Norway fall into the category we have labeled "Social Democratic welfare state." So, with some reservations, do Denmark and Great Britain.

The social insurance systems in Germany, Austria and the Benelux countries come closest to the ideal of the "conservative welfare state." Finland and France also come within this category although their systems display a number of atypical features.

The U.S.A. could be described as a fairly typical "liberal welfare state." The same, with some reservations, holds true for Australia, Japan, Canada and Switzerland.

In some countries, social welfare systems are still in their infancy. This is particularly true of Ireland, Greece, Portugal and Spain.

It remains to add that some countries are assigned to different categories by authors who have written about welfare systems. Such is the case with Great Britain, Ireland, Italy and the Netherlands (cf. Schmidt, 1988 as opposed to Wagschal, 2000).

If, with some reservations, we differentiate between conservative welfare states and states with Social Democratic, liberal and emergent systems, we can observe differences with regard to levels of welfare and development patterns. These dissimilarities suggest that—despite its shortcomings—the typology under discussion enables us to draw a number of important distinctions.

Figure 2 represents the development of social expenditure in four groups of OECD countries which have been classified in the manner just described. If we compare the time series for Germany in Figure 1b and Figure 2a, we immediately notice differences concerning the trends and levels of the two series as well as the length of the period under review. The series in Figure 2 are shorter because before 1980 we lacked data collected continuously with the aid of methods permitting valid international comparisons (so-called SOCX categories).[30] In Figure 1b the level for the 1980s is higher, and the regressive tendency is more strongly marked than in Figure 2a. These differences are due to statistical factors. Figure 1b is based on a national (German) definition of the social budget, while Figure 2a is based on a definition framed by the OECD (1998a, p. 9) with a view to facilitating international comparisons. The OECD definition reads as follows:

> "Social expenditure is the provision by public (and private) institutions of benefits to households and individuals in order to provide support during circumstances which adversely affect their welfare.
>
> Such benefits can be cash transfers, or can be the direct ("in kind") provision of goods and services, provided that the provision of the benefits constitutes neither a direct payment for a particular good or service nor an individual contract or transfer."

The SOCX categories set up on the basis of this definition comprise 13 kinds of benefits. The advantages in question are presented in Table 5, which also includes the more or less similar ESSPROSS functions considered by EUROSTAT (1996; 2000).

The comparisons in Table 5 convey a more or less accurate idea of what is hidden behind the time series represented in Figure 2. For further details the reader is referred to the relevant section of OECD (1998a, pp. 18-23).

If we compare the developments represented in Figure 2, we can observe the following phenomena:

- There are only slight differences between conservative and Social Democratic welfare states. In conservative countries, the level of the social expenditure ratio is generally between 15% and 30%, while the corresponding level in Social Democratic countries tends to range between 20% and 35%. On the whole, the regressive tendencies which become apparent in the 1990s are somewhat more strongly marked in conservative countries.
- Liberal welfare states—especially the United States—are markedly different from other countries. The differences concern both the level and

2a) Conservative Welfare States
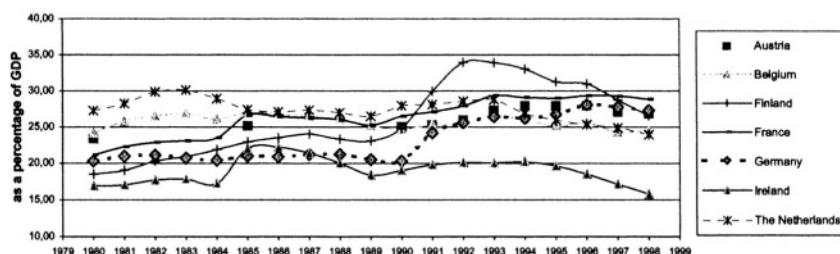
2b) Social Democratic Welfare States
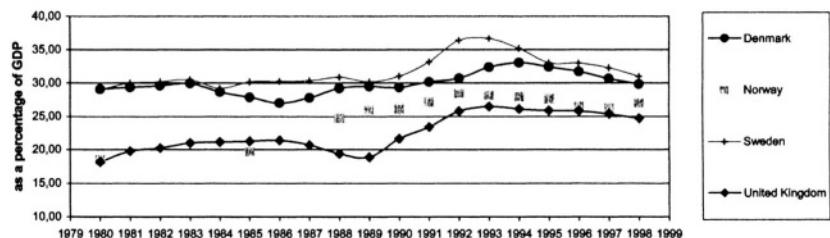
2c) Liberal Welfare States
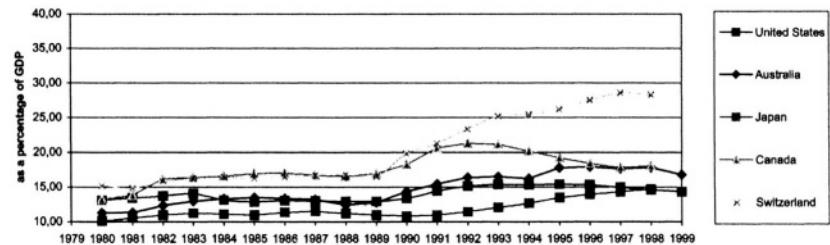
2d) Welfare States at an Early Stage of Development
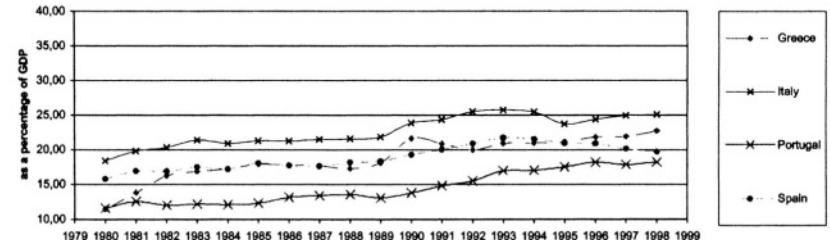
Quelle: OECD (2001) - OECD Social Expenditure Database 1980/98, Paris.

FIGURE 2. The Development of Public Social Expenditure.

TABLE 5.
SOCX Categories and ESSPROS Functions

SOCX categories	ESSPROS functions	Issues
1. Old Age Cash Benefits	Old Age	
2. Disability Cash Benefits	Disability	
3. Occupational Injury and Disease	Sickness (and Disability)	Not clearly defined in ES-SPROS
4. Sickness Benefits	Sickness	No distinction between cash benefits and services in SOCX
5. Services for Elderly and Disabled People	Old Age and Disability	
6. Survivors	Survivors	
7. Family Cash Benefits	Family	
8. Family Services	Family	
9. ALMP (Active Labour Market Programs)		Not included in ESSPROS (There is another EURO-STAT module: module PMT.)
10. Unemployment	Unemployed	Not the same coverage/definitions
11. Health	Sickness	Only public health expenditure from OECD/Health database
12. Housing Benefits	Housing	
13. Other Contingencies	Social Exclusion	Some differences in coverage

Source: OECD (1998a, p. 19).

the development of the social security systems. The level is lower, while development is more regular.
- The Mediterranean countries presented in Figure 2d set up their social insurance systems later than the other countries under discussion. With the exception of the Italian system, the social insurance schemes in the Mediterranean countries are still in their infancy. As a result, their initial level is much lower than that observed elsewhere, and they are characterized by steady growth.

If we aggregate the data for the individual groups of countries, the differences between their systems become even more apparent, and the typology described above enables us to make even more subtle distinctions. The developments that can now be ascertained are represented in Figures 3a and 3b.

The diagrams, however, are partly based on data which have been demarcated in different ways (ILO data versus OECD data), and they only reflect

3a) Social Security Benefits as a percentage of GDP 1950-1987 (ILO data)
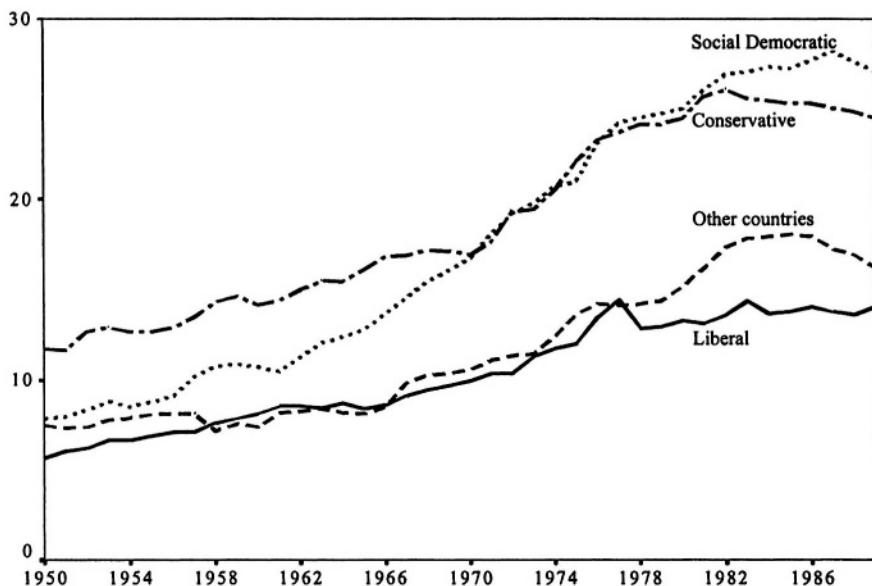
3b) Social Security Benefits as a percentage of GDP 1980-1995 (OECD data)
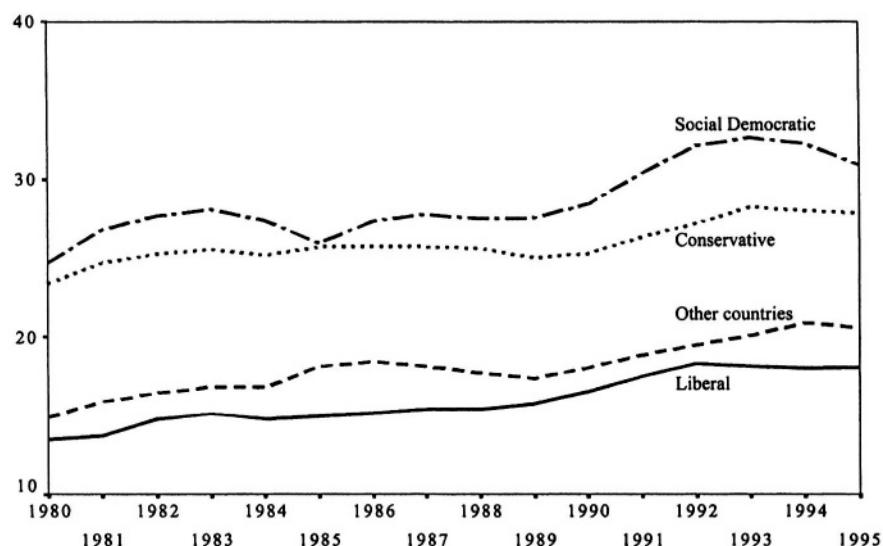
FIGURE 3. The Evolution of Social Expenditure Ratios. *Source:* Wagschal (2000, p. 42 f.).

developments that took place during certain segments of the period under re-view. There is also the further complication that the boundary lines between the various groups do not entirely coincide.

Two further points need to be made here:

- Great Britain, Ireland, Greece, Portugal and Spain have been lumped into the category "other countries."
- In the early nineties there seems to have been a temporary increase in the social expenditure ratio, and this rise seems to have occurred not only in Germany, but in other countries too. This is remarkable, but it is impos-sible to decide whether the phenomenon in question is exclusively due to economic factors.

The use of SOCX data for gross public social expenditures may distort the results of cross-section comparisons. The data reflect resource allocation at different levels of government and the use of resources from different social security funds, but they do not reflect differences between national tax systems (child tax allowances and the like). Nor do they reveal differences between regulations concerning extra-budgetary expenditures such as sick pay.[31]

An indicator has been developed for the OECD with a view to eliminating these distortions (Adema 1997). Table 6 shows how the corrections have been made, while Figure 4 represents the resultant social expenditure ratios of the countries that have been compared. Space precludes a discussion of the indi-vidual corrections made by Adema, but a glance at the table clearly demon-strates that the elimination of the distortions enables us to ascertain three im-portant facts:

- In European countries net public expenditure is much lower than gross public expenditure. In the USA, however, there is hardly any difference between the corresponding figures.
- Cross-country variations in gross public expenditure are more strongly marked than variations in net expenditure.
- Variations in net publicly mandated social expenditure are marginally higher than variations in net public social expenditure.

There are therefore good grounds for believing that the "apparently large dif-ferences in gross direct public social expenditure are due in part to institutional differences in the ways in which social objectives are pursued by governments" (Adema 1997, p. 164). Adema's assumption receives further support from the fact that in some countries educational expenses have to be borne mainly by private individuals and institutions, while in other countries such expenses are paid by the state.

All in all, there is fairly clear evidence that there *are* fundamental differ-ences between the various types of welfare states that have been distinguished

TABLE 6.

Gross and net public expenditure as a percentage of GDP (1993)

Line ± number	Item	Denmark	Germany	Netherlands	Sweden	U.K.	U.S.A.
(1)	Gross direct public social expenditure (as presently in the SOCX database)	35.2[a]	32.4	34.0	42.4	26.9	16.3
−	Direct taxes and social contributions paid on transfers	4.5	2.9	6.5	5.9	0.2	0.1
= (2)	Net cash direct public social expenditure	30.7	29.6	27.5	36.5	26.7	16.2
−	Indirect taxes on consumption purchased out of net cash transfers	4.5	3.3	2.7	4.1	2.6	0.5
= (3)	Net direct public social expenditure	26.3	26.3	24.7	32.4	24.0	15.7
+	Tax breaks for social purposes on public and private social expenditure	0.1	0.9	0.1	0.0	0.4	1.2
= (4)	Net current public social expenditure	26.4	27.2	24.9	32.4	24.4	17.0
(5)	Gross direct mandatory private social expenditure	0.7	1.8		0.7	0.3	0.5
−	Direct taxes and social contributions paid on mandatory private cash transfers	0.2	0.6		0.2	0.0	0.0
−	Indirect taxes on consumption purchased out of net mandatory private cash transfers	0.1	0.2		0.1	0.0	0.0
= (6)	Net current mandatory private social expenditure	0.4	1.0		0.4	0.2	0.5
(7)	Net current publicly mandated social expenditure [4+6]	26.7	28.2		32.8	24.6	17.5

TABLE 6.
Continued

Line ± number Item	Denmark	Germany	Netherlands	Sweden	U.K.	U.S.A.
Memorandum adjustments:						
Tax breaks on pension programmes[b]	I/A	I/A	I/A	I/A	3.1	1.0
Memorandum items:						
Pensions under administrative extension			0.7			
Indirect taxes	20.2	15.2	13.9	17.3	16.3	9.1
Related to GDP at market prices						
Gross direct public social expenditure	30.5	28.7	30.6	38.3	23.4	15.0
Net current public social expenditure	22.8	24.0	22.4	29.2	21.2	15.6
Net current publicly mandated social expenditure	23.1	24.9		29.6	21.4	16.1

[a] Values of tax breaks for social purposes for the Netherlands concern 1994.
[b] Information not available.
Source: Adema (1997, p. 163).
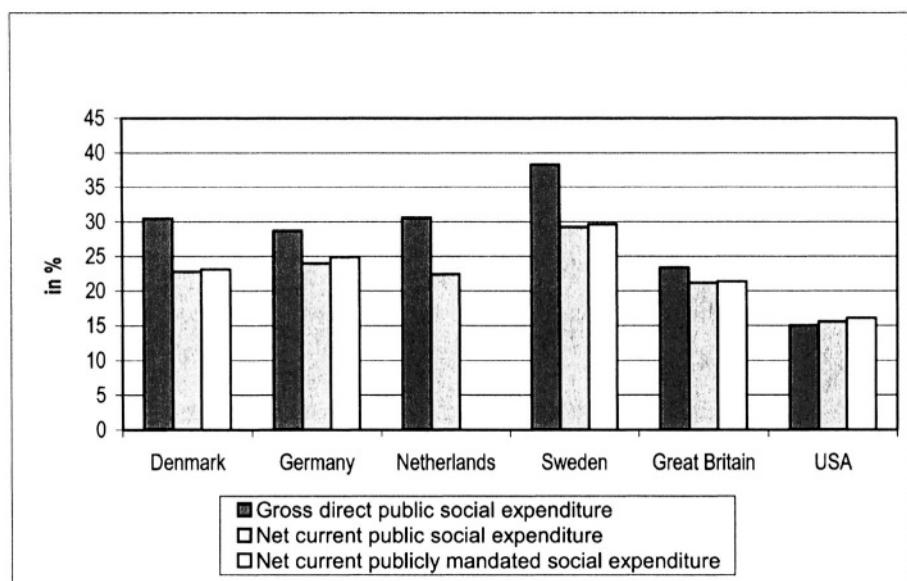
FIGURE 4. Gross and Net Social Expenditure in relation to GDP at Market Prices, 1993.

TABLE 7.

A Functional Breakdown of Social Spending in the EU (1994)

Social Risks and Facts	Country											
	B	D	DK	E	F	GB	GR	I	IRL	L	NL	P
Old Age/Surviving Dependants	44.2	41.2	33.6	42.6	43.7	41.3	68.8	64.0	27.5	46.0	36.8	40.1
Sickness/Disablement/ Industrial Accident	35.4	38.8	27.0	36.2	34.0	31.3	24.5	29.9	36.8	38.0	43.6	47.9
Family/Maternity	8.1	7.6	11.6	1.7	9.6	11.4	1.2	3.6	13.1	13.6	5.4	5.2
Unemployment/Labor Market	11.0	9.2	16.8	18.1	8.1	7.3	2.7	2.5	17.2	2.3	10.4	5.8
Miscellaneous	1.3	8.0	1.0	1.4	4.6	8.7	2.8	0.0	5.4	0.1	3.8	1.0
Total	100	100	100	100	100	100	100	100	100	100	100	100

Source: Bundesministerium für Arbeit und Sozialordnung (1998, p. 12).

B = Belgium; D = Germany; DK = Denmark; E = Spain; F = France; GB = Great Britain; GR = Greece; I = Italy; IRL = Ireland; L = Luxembourg; NL = Netherlands; P = Portugal.

by political scientists. Contrary to expectations, however, the differences between the social expenditure ratios of conservative and Social Democratic welfare states are relatively slight.

A similar picture emerges if, instead of comparing systems in their entirety, we examine the functions assumed by various kinds of benefits in major welfare states. The results of such an investigation are shown in Table 7. A glance at the figures shows that although there are marked differences between EU member states, the share of funding that goes on individual welfare benefits is broadly similar throughout the EU. The only exceptions are Greece and Italy, where spending on pensions is much higher than in other EU countries.

Present-day social insurance systems are similar with respect to benefits as well as financing, and the similarities between their financing methods are becoming increasingly evident. Table 8 shows that this is particularly true of European systems, which are all supported by three pillars: employees' contributions, employers' contributions, and state subsidies. Employees' contributions range between 10.1% in Denmark and 47.8% in the Netherlands; employers' contributions vary between 9.2% in Denmark and 50.3% in Spain, and government subsidies range between 16.3% in the Netherlands and 75.6% in Denmark. Only the Danish and Irish schemes are mainly financed by taxes, which are supplemented by other kinds of revenue (chiefly interest earnings).

Despite existing disparities, Europe's welfare states have moved much closer together with respect to levels of welfare, development and structures. This harmonization process is apparently still under way, for the systems that

TABLE 8.
The Financing of Social Security Benefits in the EU (1994)

Country	Employees' Contributions (incl. contributions made by pensioners and self- employed people)	Employers' Contributions (incl. presumed contributions for civil servants)	State Subsidies	Miscellaneous (e.g., interest earnings)
B	26.5	43.1	20.6	9.8
D	32.0	38.4	26.7	2.9
DK	10.1	9.2	75.6	5.1
E	18.0	50.3	29.6	2.1
F	27.4	49.1	21.5	2.0
GB	15.6	26.1	43.9	14.4
GR	26.7	45.7	19.2	8.4
I	16.3	46.5	35.0	2.2
IRL	15.2	23.0	61.0	0.8
L	21.4	29.5	42.6	6.5
NL	47.8	20.2	16.3	15.7
P	21.3	34.4	37.4	6.9

Source: Bundesministerium für Arbeit und Sozialordnung (1998, p. 14).
B = Belgium; D = Germany; DK = Denmark; E = Spain; F = France; GB = Great Britain; GR = Greece;
I = Italy; IRL = Ireland; L = Luxembourg; NL = Netherlands; P = Portugal.

are mainly financed by contributions display "a tendency to rely increasingly on taxes rather than contributions" (European Commission, 1996, p. 81).

As we have already said, space precludes detailed treatment of the specific features of all the major state social insurance systems. We shall therefore content ourselves with a few bibliographical references:

- MISSOC (2001) describes the organizational structures of social insurance institutions in the EU and the EEA. There is also a great deal of information about the way these institutions distribute various kinds of welfare benefits.
- Germany's Federal Ministry for Labor and Social Order (Bundesministerium für Arbeit und Sozialordnung, 1998) has published a study of the various types of benefits available to people living in the EU states. In this work, the primary focus is on comparative law.
- Siegel (2000) has researched the causes behind the developments represented in Figure 3; his study is based on multi-dimensional regression analyses and takes account of political factors.
- Finally, comparative studies of various countries have been presented by Börsch-Supan and Miegel (2001).

2.3.   Social insurance and the European Single Market

Tensions may develop between national insurance systems within the EU.
The reasons for this are relatively simple. The Western Europeans have created
a single market (EEC Treaty, Article 14) governed by laws that guarantee the
free movement of citizens (Articles 39-48), goods (Articles 28-30), services
(Articles 49-55), and capital (Articles 56-60). Yet responsibility for social pol-
icy still rests with individual member states, whose governments are merely
obliged to respect the aforementioned rights to free movement (EEC Treaty,
Article 42).

A further complicating factor is the eastward expansion of the EU, which
will probably result in an increase in the number of immigrants. What is more,
the introduction of EU citizenship may make it easier for immigrants to claim
welfare benefits in the EU states in which they have settled.

Three examples will serve to illustrate the problems we have in mind:

A worker living in a country outside the EU may require a transplant which
his national health service refuses to pay for, so he may move to an EU state
where his transplant operation will be free of charge.

- In the wake of the eastward expansion of the EU, an Eastern European
  worker might sign a bogus fixed-term work contract in one of the EU
  15 countries, thereby acquiring rights to welfare benefits such as income
  support. If he remains in his home country but pretends to be permanently
  resident in the state where he signed his fake contract, his welfare benefits
  will give him considerable purchasing power.
- Our third example is of a more general nature. A besetting economic is-
  sue is raised by the relationship between (1) national responsibility for
  social policy and (2) EU citizenship and the four basic freedoms referred
  to above. How can we preserve the advantages attached to our basic free-
  doms without allowing our welfare systems to be exploited by unscrupu-
  lous people?

The explosive nature of these problems becomes glaringly obvious when
we examine the legal situation and consider the various ways in which our
social security systems can be abused.

According to the European Court of Justice, there are only four conditions
under which restrictions can be imposed on the aforementioned basic free-
doms: "They [*sc.* the restrictions] must be justified by imperative requirements
in the general interest; they must be suitable for securing the attainment of the
objective which they pursue; and they must not go beyond what is necessary
in order to attain it."[32]

This raises a question of fundamental importance. Population shifts can
sometimes be described as "productive" because they move factors to places

where productivity is higher or because they transfer goods to places where they will be more useful. In EU member states such population shifts should not be hampered by legal constraints entailing cuts to social security benefits.

"Unproductive" population shifts, however, are an entirely different matter. Since such shifts are merely designed to exploit foreign social security systems, some measures ought to be taken in order to prevent—or at least hamper—this kind of migration. We can put forward two arguments in support of such preventive measures:

- Most Europeans still lack a strong sense of solidarity with citizens from other EU member states.
- There are still enormous disparities between the social security systems of EU member states. The German and Greek systems, for example, are in entirely different stages of development. The Greeks are not yet in a position to provide uniform minimum standards of social protection for EU-countries, while Germans would consider such minimum standards unreasonably low.

In the EU, "productive population shifts" are facilitated by ordinance 1408/71 regulating the social security of migrant workers. This regulation is supplemented by implementing ordinance 574/72, which is concerned with classic types of benefits and is based on the following principles:[33]

- A worker is subject to the laws of only one country. As a rule, this country is the one in which the worker is employed (country of employment principle).
- EU-citizens are to be treated equally (equality of treatment).
- A welfare beneficiary is not normally obliged to reside in the country from which he receives his benefits (waiving of residence clauses).
- In order to ensure that a worker who moves from one EU-country to another will not have to begin a qualifying period all over again, the relevant periods are added together (aggregation of periods of insurance, employment or residence).
- If an immigrant claims an old-age pension or a surviving dependents' pension, each of the states concerned will calculate the amount of the benefits that would have had to be paid if the claimant had spent all the relevant periods in the relevant country. The sum which an EU state has to pay will depend on the length of the periods during which the claimant was resident of the relevant state (pro rata determination of benefits).

In order to hamper migration intended to exploit the social safety net, the Scientific Advisory Board of the German Finance Ministry (Wissenschaftlicher Beirat beim Bundesministerium der Finanzen, 2000, pp. 34-94) has proposed

what is known as "delayed integration." When someone takes up a post out-side his home country, the "country of employment principle"[34] may not be applied immediately. The immigrant worker may have to wait for about five years before he is integrated into the social security system of the employment state. This kind of regulation would not fully safeguard against "hit-and-run actions" designed to exploit foreign social security systems, but it would make such actions much more difficult.

This brings us to the end of our survey of modern social insurance systems and special problems in the EU. We may now turn to a consideration of the economic effects of these schemes.

3.   THE ECONOMIC EFFECTS OF SOCIAL INSURANCE: A GENERAL OVERVIEW

3.1.   The Effects on Income Distribution and Economic Stability

To classify the economic effects of social insurance we have to ask two fun-damental questions. First, do these effects concern distribution, stabilization or allocation? Second, do they correct market failures or cause distortions?

Let us begin by considering the problems associated with the distribution of welfare benefits. Who are the beneficiaries, and who pays for the benefits? The answers to these questions will depend to a large extent on the answers to two other questions, both of which are basically concerned with allocation. First, are there any gaps in the insurance markets that play a crucial role in distributing income within the economic systems under discussion? Second, how well do these insurance markets function?

In addition, we must consider whether social policy can be used to correct market failure in insurance markets which assume a prominent role in distrib-uting income. This might indeed be the case if the state could use public law to reduce transaction costs which would prevent the emergence of the kinds of insurances in question.

As we have already explained, market failure might well be brought about by those typical insurance risks which are described as moral hazard, adverse selection and risk cumulation, and the problems posed by such risks are often aggravated by the fact that many people adopt a short-sighted approach to the problems associated with insurance.[35]

Since health[36] and unemployment insurance (see Table 7) are particularly prone to such forms of market failure, it might be worthwhile to devise a com-pletely new type of policy which would borrow only a few of the characteristic elements of classic welfare policies, and which would boil down to a sort of conscious social risk management (Holzmann and Jørgensen, 2001).

Owing to mismanagement and insufficient membership, health insurance companies have often gone bankrupt in the 19th century (see Section 1.2 (p. 3)). The fatal weaknesses of such enterprises can probably be explained by two factors. First, the problems posed by moral hazard could only be resolved in local communities. Second, people often remained uninsured although in the long run it would have been in their best interest to take out an insurance policy.

As far as moral hazard is concerned, state-run health insurance companies are just as vulnerable as their private counterparts. Typical instruments used by insurance companies (deductibles, qualifying periods and inspections) can be employed on the basis of private law, and in most cases they ought to be sufficient. The problems posed by a short-sighted approach to insurance can be resolved by legislative measures imposing compulsory insurance. However, there is no possible justification for compulsory state insurances or national health services. Indeed, there are good grounds for believing that such institutions do more harm than good.[37]

Owing to the obvious problems posed by moral hazard, adverse selection and risk cumulation, our ancestors never tried to set up a private unemployment insurance system in the modern sense. In continental Europe, however, there were some organizations which fulfilled similar functions, and which might be considered as the forerunners of our present-day institutions. The organizations in question were generally supported by labor unions, their influence was limited to a fairly confined geographical area, and they doled out welfare payments in an attempt to prevent the exploitation of unemployed workers. Following the example of the city of Ghent, local authorities often supported these organizations in the hope of economizing on support payments (Frerich and Frey, 1996b, p. 153; Hertner, 2002, p. 32 f.). The main obstacle to the emergence of a genuine unemployment insurance system—the tendency to adverse selection—was only removed when the state introduced compulsory insurance. Many countries modeled their systems on the British scheme, which assumed a pioneering role in this domain (Hennock, 1982).

There are good grounds for believing that moral hazard, adverse selection and risk cumulation have had little influence on old-age insurance systems, for life insurance schemes were set up on a purely private basis at a relatively early date (Schewe, 2000). Scholars have never been able to decide whether life insurance policies can be considered as merit goods, though one might argue that life insurance can indeed be classed as a merit good because many people tend to prioritize immediate consumption in their salad days, and it is only late in life—in many cases too late—that they recognize the need to make adequate pension provisions.

The improvidence displayed by such people might well provide a justification for measures which would ensure a minimum level of social protection

by obliging everyone to make provision for their old age. Recent opinion polls suggest that—at least in Germany—most citizens believe that compulsory insurance is necessary. Over 70% of those surveyed advocate a reform of the old-age-pension system involving a reinforcement of measures which oblige people to take out private insurance in order to ensure that they have enough to live on when they retire. By contrast, only slightly more than 40% of respondents are in favour of a reform which would not involve any coercive measures (Mannheim Research Institute for the Economics of Aging, 2002).

Since we lack a suitable yardstick against which to measure our data, it is virtually impossible to discover how the aforementioned main types of insurance influence the distribution of income within a country.

There is pretty general agreement that social insurance systems, together with poor relief measures, have succeeded in eliminating the worst forms of poverty. There is also broad agreement that these systems and measures have provided a kind of insurance against physical handicaps as well as the disadvantages suffered by people who come from a deprived working-class background. There is, however, scope for considerable disagreement on costs[38] and the overall effects on the distribution of income; and it is an open question whether social insurance systems merely serve to correct market failures, or whether interest groups who play an active role in this domain create new types of distortion.

Such distortions are less likely to occur in systems based on the Beveridge model than in those based on the Bismarck model. As we have already pointed out, systems based on the Beveridge model tend to conform to the welfare and provision principles. In Bismarckian systems, by contrast, welfare authorities often make payments which, strictly speaking, have little to do with insurance in the private sector.

It is, moreover, important to remember that there is a phenomenon known as "expenditure snatching," which presents certain points of resemblance to tax shifting (Zimmermann and Henke, 2001, pp. 265-270). Nonetheless, some economists incline to the view that, all things considered, the distribution effects in states with traditional social security systems are positive (cf. Auerbach and Kotlikoff, 1987; Diamond, 1977; Kotlikoff, 1998; Rothschild and Stiglitz, 1976), The ideas put forward by authors like Auerbach and Kotlikoff are perfectly compatible with the findings reported in an empirical study by Caminada and Goudswaard (2001), who arrive at the conclusion that recent measures designed to dismantle the social security systems in the OECD states (e.g., in Great Britain and the Netherlands) have led to greater inequalities in the distribution of disposable income.

The distribution effects produced by social insurance are not only intragenerational. There are also intergenerational effects which are mainly due to old-age insurance. Most economists take the line that when a pay-as-you-go

system is introduced, the first generation of pensioners will get a better deal than later generations.[39] The only exception is a situation where there is reason to assume a strong inheritance motive which can trigger private reactions likely to offset government measures (cf. Barro, 1974).

In addition to the redistribution effects we have assumed, we can observe effects produced by changes in the structure of the population. These changes have triggered reforms or created situations in which people have become aware of the need to institute reforms. These issues will be discussed in section 4. In this section we simply wish to point out that generational accounting has been used in order to quantify the effects that can be seen in many countries (Auerbach et al., 1999). In most cases there is a tendency to burden future generations, and many people believe that our descendants will be overburdened.

While the distribution effects of social insurance systems are considered to be partly harmful, the stabilizing effects of these schemes are generally held to be beneficial. By stabilizing effects we mean a curbing of cyclical fluctuations in price levels, the unemployment rate, the growth rate, and the current account balance.

Social insurance payments are subject to countercyclical variations, and this is particularly true of unemployment insurance payments. In conjunction with variations in tax revenue, they therefore act as automatic stabilizers, which work without delays caused by forecasting or decision-making problems.

The European Growth and Stability Pact, which takes full account of the stabilizing effects produced by social security systems, was designed to ensure that participating countries follow a steady financial policy. The signatories of the Pact are therefore required to maintain balanced budgets in the medium term, and if they keep their budgets in order, they can make the most of the scope offered by the net borrowing limit (3 per cent of GDP), thereby allowing the automatic stabilizers to take effect.

It is not only the overall economic situation that is stabilized by social insurance. A stabilizing effect can also be observed in regions threatened by structural crises. Thus, if a monostructural region is facing a crisis that jeopardizes the industry where most people work, a social insurance scheme will normally preserve enough purchasing power to cushion the impact of the crisis on consumer-oriented sectors such as local handicrafts, retail trade and the catering industry.

In the wake of German reunification this stabilizing effect played a major role. Following the introduction of the Deutschmark and free trade, the entire industrial system of the former GDR collapsed because it was not sufficiently competitive. However, since all the citizens of the new Länder were integrated into the West German social insurance system, their purchasing power was stabilized and East Germany's industrial sector was offered a new lease of life.

3.2.   Allocative effects: efficiency, labor market distortions, and capital market distortions

When we investigate the allocative effects of social security systems, we have to consider how social policy affects the use of resources and aggregate welfare. All social security systems produce three kinds of allocative effects:

- general excess burden effects connected with the financing of the schemes,
- additional special effects on the labor market, and
- additional special effects on the capital market.

Special attention must be given to the problems posed by compulsory insurance, for there are a great many countries where people who are obliged to take out insurance[40] are not free to choose their insurer. They may have no choice but to apply to the state or an institution that is dependent on the state, and this kind of institution may even assume the role of a service provider.[41]

In such cases "make or buy" questions arise, and these questions are linked to issues concerning the privatization of social security systems (Diamond, 1999). In the present study we can only touch on these subjects.

Economists have no difficulty in identifying general excess burden effects in tax-financed systems. Since the taxes in question are not lump sum taxes, they produce unwelcome substitution effects that result in welfare losses.

Such effects, however, may also occur in contributory schemes. Employers consider employers' contributions as a payroll tax that puts up the cost of the labor factor. By the same token, employees regard employees' contributions as a sort of tax when the equilibrium between work and pay is disturbed by redistributive elements inherent in the system.

Together with other work-related charges, employers' and employees' contributions form a kind of wedge which widens the gap between gross hourly pay[42] and net hourly pay[43]. As a result, employers and employees base their calculations on different prices. Together with some kinds of burden shifting this distorts the coordination of production and consumption decisions. As a result, in many markets welfare losses are incurred.

Such phenomena are particularly likely to occur in situations where employees do not consider the two contributions in question as fair insurance premiums which they would have been prepared to pay of their own accord. Owing to a large number of redistributive elements inherent in real contributory schemes, it is, therefore, unlikely that employees would ever agree to pay contributions on a voluntary basis.

At this point it may not be amiss to draw attention to a further complicating factor. The formal distinction between employers' contributions and employees' contributions deludes employees into underestimating the cost of social insurance.

The odds are that general excess burden effects will be lower in tax-financed systems than in contributory schemes. In tax-financed systems the basis of assessment is somewhat broader and tax rates therefore somewhat lower, so that the excess burden tends to be reduced (Hopmann, 1998, p. 281).

In order to give a full description of general excess burden effects, we would have to use a computational general equilibrium model. Using partial analysis, Rosner (2003) has, however, provided a qualitative description of the major knock-on effects.

Mention should also be made of two studies by Homburg (1990) and Homburg and Richter (1990). According to these authors, excess burden effects—especially in pay-as-you-go pension schemes—are so strong that transitional generations might be able to avoid most double burden effects by means of an adroit reduction of excess burden effects during a transition from a pay-as-you-go system to a funded scheme.

Additional effects on the labor market can be observed when social insurance (mostly in conjunction with poor relief measures) functions like a minimum wage, when there are incentives for people to take early retirement, or when idleness is subsidized. The additional effects in question reinforce general substitution effects in the labor market. To be more precise, income tax and income-related social insurance contributions create a situation in which many people prefer to work less and have more leisure time.

A minimum wage above equilibrium price will result in classic unemployment. Similar effects may be produced by the policies pursued by labor unions, by the interaction between such policies and social measures taken by the government, or by a situation where the level of social insurance implicitly defines a minimum wage level.

Legislation providing effective protection against unfair dismissal will create strong incentives for people to take early retirement. To take just one example, an employer may find that it makes economic sense to pay a company pension to an early retiree who is drawing unemployment benefit. Economists have observed this kind of phenomenon in countries like Germany, and they have also drawn attention to the fact that benefits are not adjusted in an actuarially fair manner when an employee takes early retirement (Börsch-Supan, 2001a, p.21).

Generous child benefits may also give people a powerful incentive to withdraw—at least temporarily—from the labor market. This is particularly the case with women.

Invidious effects are often produced by the means testing of income support and old-age benefits. When income support is set off against earned income, many people are subject to a sort of confiscatory taxation, and workers whose productivity is low often choose to give up badly paid jobs and live on in-

come support. In such cases the government might be accused of subsidizing idleness.

This problem could be resolved by means of regulations which would have more or less the same effect as the introduction of a negative income tax, and which would assign separate resources to the attainment of two goals: (1) to ensure that those who are able to work are employed at productivity related rate, (2) to guarantee a minimum level of income. In many cases, however, politicians shy away from taking appropriate measures because extremely high transaction costs would have to be defrayed in order to ensure that government money is not paid out to the wrong people. It remains to add that the Americans are currently carrying out a promising experiment designed to provide a workable solution to such problems. The experiment is known as the Wisconsin Works Program[44].

It is very hard to assess the quantitative significance of the allocative distortions caused by financing and incentive effects in the labor market. In order to make such an assessment, we have to explain exactly what we mean by the "allocatively neutral organization of a social security system."

This raises a vast array of questions such as the following:

- How should we deal with benefits related to retirement or the education of children? Are these benefits overall economic costs associated with the labor factor? Should we consider a minimum of cover as something for which society at large is responsible?
- Should we ask what a society would look like if cover for the retired and for child care had to be provided on an individual basis by private insurance companies, while the government would merely guarantee the existence of appropriate compulsory insurance or insurance firms?
- How should actuarially fair risks be ascertained when insurance claims are filed?
- What kind of data ought to be subject to data protection laws?

Up to now, nobody has managed to devise an overarching solution to these problems. Hopmann (1998), however, suggests an approach which is worth discussing. He frames a definition of allocative neutrality, and on the basis of this definition he makes international comparisons which reveal considerable quantitative differences in the degrees of distortion that can be observed in various countries. In Denmark, Great Britain and the Netherlands, for instance, social insurance contributions appear to be under relatively firm control, but in Belgium and Germany they have sent labor costs soaring (Hopmann, 1998, p. 281).

On the demand side, additional effects on the capital market are due to the fact that social security systems lead to a higher degree of specialization and are reflected in the risk structures of real and human capital investments. So

far, however, no one has put forward a full-blown theory to explain these facts, and no attempts have been made to assess the quantitative significance of the effects in question. Metze, Prinz and Lübke (2003) present a first attempt at simulating the impact of aging on capital markets.

On the supply side, social security systems influence the level and structure of the supply of monetary capital in the global capital market. A great deal depends on whether pension insurance is financed by a fully funded scheme[45] or a pay-as-you-go system[46].

We might draw a parallel between these pension systems and private life insurance schemes. When a life insurance scheme is set up, policyholders have to save up a considerable amount of money before they can collect pension benefits. Initially, therefore, more money is paid in than is paid out, and in the course of time a stock of capital is built. At a later stage, once membership has levelled out, this capital stock may be preserved even without interest rate effects, provided that outpayments can be financed by current inpayments. In a pay-as-you-go system, by contrast, inpayments are immediately used to cover expenses, so that it is virtually impossible to build a substantial stock of capital.

However, what is true of an individual enterprise is not necessarily true of the economy as a whole. Let us suppose that the inhabitants of a given country make provision for their old age by setting aside a portion of their current income. If compulsory insurance is introduced, it will make little difference whether the powers that be opt for a fully funded scheme or a pay-as-you-go system. In both cases, the odds are that current private savings will be reduced. Indeed, under certain circumstances, the reduction in savings brought about by the two systems may even be identical (Homburg, 1988, p. 57).

In a fully funded scheme, however, the reduction in private savings is off-set by the constitution of a capital stock with the insurer. In a pay-as-you-go system, by contrast, this is by no means the case (Homburg, 1988, p. 28).

People often infer from this that the essential difference between the two systems has to do with the level of an economy's capital resources and the way in which this affects growth.[47] This inference, however, is not based on solid empirical evidence. There are two caveats to be made:

- Our first caveat has to do with the savings amassed and invested by pensioners before the introduction of compulsory insurance. If a pay-as-you-go system is set up, these savings are no longer required to cover retirees' living expenses. Do pensioners spend or bequeath this money? If the pensioner is an out-and-out egoist, he may simply go on a spending spree, but he might prefer to bequeath his savings if he is inclined to exercise thrift and is anxious to provide for his children.[48]
- Our second caveat has to do with costs that may have to be borne by the taxpayer after the introduction of a fully funded scheme. Additional tax

money may be required to support senior citizens who have not made adequate private provision for retirement, or whose savings have been wiped out by war or inflation.

In reality, therefore, the difference between the two systems is probably less salient than appears at first glance.

When we compare and contrast fully funded schemes and pay-as-you-go systems, it is not sufficient to consider problems associated with capital formation. Special attention must also be given to differences regarding proneness to certain risks.

The pay-as-you-go system will be in jeopardy if there is a fall-off in the number of contributors. One might, for instance, imagine a worst case scenario where an entire generation would remain childless. The pensioners who belong to this generation would have nobody to support them in their old age, and the system will not be able to meet the obligations on the basis of a generational contract.[49] The generational contract will collapse like a Ponzi debt.

Fully funded schemes are less dependent on demographic developments because capital can be invested abroad and a nation's last generation could—in theory at least—live on pensions funded by foreign investments.

Nonetheless, fully funded schemes are prone to three risks:

- The first risk is one that is typically associated with capital investments. Good examples are provided by war damage and long-term speculative bubbles in share markets.
- The second risk is a concentration of power in a small number of pension funds.
- The third risk is dependence on the age structure of the population. This kind of risk must be taken very seriously if we accept the validity of the life cycle savings accumulation model and assume that a world-wide aging process will be triggered by factors beyond our control. If this kind of process were to set in, senior citizens all over the world would try to realize their assets, thereby perturbing the circular flow of money. Securities would be sold at a loss, and deflation would ensue.

Since it is impossible to assess the extent of such risks, it is hardly surprising that many people advocate a mixed system (Börsch-Supan, 2001b, p. 10 f.). A persuasive argument in favor of this kind of scheme is that the pre-funding of liabilities could be organized in such a way that it would be fertility-dependent (Lassila and Valkonen, 2001). Some economists add that fully funded schemes are preferable for two reasons. First, they are more efficient and might lead to greater prosperity. Second, they would be less prone to misuse by politicians if their underlying principles were enshrined in the constitution.[50]

4.   POPULATION AGING AND PROPOSALS FOR PENSION REFORM

In all the advanced industrial societies, populations are aging fast. As a result, many of these countries' pension programs have excited violent discussion and been analyzed in internationally oriented studies.[51] A collaborative volume edited by Börsch-Supan and Miegel (2001) offers an informative survey of six major industrial states. Separate chapters have been devoted to Chile (Schmidt-Hebbel), Germany (Börsch-Supan), Great Britain (Disney), the Netherlands (Bovenberg and Maidam), Switzerland (Brombacher-Steiner), and the United States (Wise).

All these countries' pension systems were set up before the Second World War and underwent vigorous expansion in the sixties and seventies. Since the eighties they have been trying to cope with the problems posed by a cost explosion.

These problems have sparked a tense welfare-reform debate. All the countries concerned have begun carrying out reforms, but Germany is lagging behind other advanced industrial nations although the German system is most immediately threatened by future demographic developments.[52]

A glance at Table 9 shows striking parallels as well as considerable differences between old-age pension schemes in various countries. These similarities and dissimilarities are strongly reminiscent of our findings concerning social security systems in general.

As we can see from the descriptions in Table 9a, there are appreciable points of divergence between institutional structures. A similar impression is conveyed by Table 9b, which represents the relative importance of approximately comparable subcomponents of the old-age pension schemes. As the third pillar grows in importance, so does the share of financing for fully funded schemes.

At the same time we can observe certain points of resemblance. The first of these is the fact that each of the old-age pension systems can be divided into three sub-systems (the "pillars"):

- The first pillar consists of a sub-system which relies on compulsory financing. It provides at least a basic safeguard, but often offers more than this. We note the predominance of defined benefits and pay-as-you-go funding.
- The second pillar is constituted by employer-related pension plans, which are generally subsidized by tax breaks and based—at least in part—on fully funded schemes.
- The third pillar is made up of private savings. As a rule, these savings are also subsidized, provided that they take the form of long-term investments. In most cases they constitute fully funded systems on the basis of defined contributions.

TABLE 9.
The Three Pillar Scheme

(9a) The Three Pillars

	Germany	Netherlands	Switzerland	Great Britain	United States	Chile
First Pillar	Public Retirement Insurance (GRV)[a]	Public basic pension (AOW)[b]	Public basic pension (AHV/IV)[c]	Public basic pension (Part of NI)[d]	Social Security	Public minimum pension
Second Pillar	Company pensions (mostly reserve accounts)	Employer-related pensions (mostly defined-benefit plans[e])	Employer-related pensions (BVG[f])	Public earnings-related scheme (SERPS[g]); company pensions; private pensions; stakeholder pensions	Company pensions (defined-benefit and defined-contribution[h] plans; 401(k) plans[i])	Pension Funds (AFP)[j]
Third Pillar	Own savings (mainly whole life insurance), other income	Own savings (mainly pensions funds and whole life insurance), other income	Own savings (mainly whole life insurance and pension funds), other income	Own savings (mainly pension funds, supplements to company pensions), other income	Own savings (mainly IRA[i] and Keogh[i]), other income	Own savings (mainly pension funds, supplements to company pensions), other income

TABLE 9.
Continued

(9b) Share of Income from each Pillar (in %)

	Germany	Netherlands	Switzerland	Great Britain	United States
First Pillar	85%	50%	42%	65%	45%
Second Pillar	5%	40%	32%	25%	13%
Third Pillar	10%	10%	26%	10%	42% (of which 25% is earnings)

(9c) Replacement Rate of Retirement Income

Germany	Netherlands	Great Britain	United States	Switzerland
85%	86%	72%	85%	80%

[a] GRV = Gesetzliche Rentenversicherung (a mandatory pay-as-you-go pension system).

[b] AOW = Algemene Ouderdoms Wet (a mandatory pay-as-you-go pension system).

[c] AHV/IV = Alters-, Hinterbliebenen und Invalidenversicherung (a mandatory pay-as-you-go pension system).

[d] NI = National Insurance (Part one is a mandatory basic pensions system, financed largely from payroll taxes on a pay-as-you-go basis).

[e] A defined-benefit plan specifies the payout that an employee is to receive according to factors such as salary and years worked.

[f] BVG = Gesetz über die berufliche Vorsorge (a mainly compulsory funded pension scheme).

[g] SERPS = State earnings-related pension system (a second pillar with an opt-out opportunity).

[h] A defined-contribution plan specifies the payments that a firm must make into a pension fund, which then invests them to earn a return for future retirees.

[i] IRAs (Individual Retirement Accounts), 401(k) plans and Keogh plans are employer-sponsored personal savings plans that offer tax advantages.

[j] AFP = Administratoras de Fondos Pensiones (Chilenian pension funds which administer compulsory contributions).

Source : Adapted from Börsch-Supan (2001b, pp. 3-5).

The second point of resemblance between the various old-age pension systems can be seen in Table 9c. In most of the countries under discussion, the replacement rate of retirement income[53] is between 80% and 86%. In Great Britain, however, it is only 72%.

It is important to note that similar or even identical replacement rates may conceal important differences which will no doubt become apparent in the course of time. Consider, for instance, these rates in Germany and Switzerland. The German rate is 85%, and the Swiss rate 80%.

In the long term, however, the Swiss rate is more likely to remain stable than the German rate. This is because the share of pre-funding is much higher in Switzerland than in Germany.

What lessons can we learn from a comparison of various systems? Börsch-Supan (2001b, p. 11) answers this question as follows:

> "Much that has been rejected as 'politically impossible' in the reform debate of one country is everyday practice for the neighbours."

This statement should be pondered by people who have misgivings about the political feasibility of reform proposals.

A wide range of proposals have been put forward by those who have taken part in welfare-reform debates in various countries. The proposals fall into three main categories:

- proposals for a parametric reform in the pay-as-you-go pillar
- proposals for a gradual transfer of weight from one pillar to another (reinforcing the financing of fully funded schemes via the second or third pillar)
- proposals for a sweeping reform of the entire system (e.g., a transition to a basic, low-level, tax-financed safeguard in the first pillar, in conjunction with a transfer of weight to the financing of fully funded schemes via the second or third pillar).

Four parametric reform options in the first pillar are currently under consideration: (1) raising contribution or tax rates, (2) lowering pension levels, (3) raising the number of taxpayers by facilitating the immigration of young, well trained workers, and (4) raising the age of retirement in order to increase the number of years spent in the labor force.

Raising retirement age is no doubt the most promising of these options. However, if this type of reform is to be successful, it will have to be combined with two other measures. First, governments will have to create more opportunities for part-time work. Second, going rates will have to be lowered if there is a fall-off in the productivity of individual workers.

It is, however, essential that lower income in the years immediately preceding retirement should not result in lower pensions. In order to avoid giving

certain people the wrong kind of incentive, the government ought to take measures to ensure that even lower retirement earnings will lead to higher pension plus earning income.

It is virtually impossible for politicians to push through sweeping reforms because such measures inevitably give rise to serious transitional problems and generally prove deeply unpopular with the median voter. But some piecemeal reforms have already been carried out, and further measures are currently under consideration everywhere.

5.   CONCLUDING REMARKS

In this article we have attempted to answer six main questions:

- What is understood by the term *social insurance*?
- Why have social insurance systems been set up?
- What do modern social insurance systems look like?
- How do these systems work?
- What kinds of malfunctions can we observe?
- What reforms have been proposed?

The term *social insurance* has extremely positive connotations. The adjective *social* carries implications of moral goodness, while the noun *insurance* indicates the presence of a safety net intended for people who, through no fault of their own, might otherwise find themselves in straitened circumstances.

However, we should not overlook the drawbacks of social insurance systems:

- The reasons for which social insurance schemes have been instituted are not purely altruistic. Indeed, some elements of these systems have been devised with positively Machiavellian cunning.
- For purely electoral reasons, politicians therefore may shy away from carrying out necessary reforms. Indeed, reforms are sometimes delayed until acute social crises erupt.
- The resources of social insurance schemes may be overstretched by demographic developments and over-generous social welfare programs.
- Social safety nets may be misused as hammocks.

In modern countries social insurance systems have assumed gigantic dimensions. One might therefore well imagine that all their major problems and effects on the economy have already been thoroughly researched. In actual fact, however, almost the opposite is the case, for a great many questions remain unanswered or have only been partially answered:

- What benefits ought to be provided by a good social security system?
- To what extent can we make generalizations on the basis of value judgments like those that are implicit in Dixon's social insurance ranking system (Dixon, 1999)?
- What kinds of challenges should a good system be able to meet?
- How well have various national systems stood the test of time?
- How far have existing systems succeeded in redistributing wealth, and how high are the efficiency costs occasioned by this redistribution?
- Given this situation, can anyone claim to be in a position to offer politicians advice based on sound academic research?
- And how should we reply to questions about ways and means of remedying the deficiencies of existing social security systems?
- What kinds of constitutional constraints would have to be made in order to prevent political misuse of the moral sentiment motive and correct serious defects due to inefficiency and insufficient insurance cover?

Unfortunately, economists who attempt to answer these questions have no choice but to rely on intuition and on simplistic models which are not amenable to empirical tests. Thus, Mirrlees' lecture on the economics of the welfare state ends with a comment which shows his unshakeable belief in instinctive knowledge and simplified representations of complex systems (Mirrlees, 1995, p. 396).

In this context, it may not be amiss to quote from the conclusion of Barr's essay on the welfare state:

> "Not only social insurance, narrowly defined, but also 'universal' benefits and social assistance are a form of insurance. By offering cover prior to birth, the welfare state is acting like ex ante actuarial insurance with a long time horizon" (Barr, 1992, p. 795).

Among other things, the present article has shown that there is no such thing as *the* welfare state. Let us hope that Mirrlees' faith in economic models is not entirely groundless, and that Barr's statement is broadly applicable to existing social insurance systems.

NOTES

1. In Germany the *Grundgesetz* (Basic Law) functions as a key part of the government and state structure.
2. The literature on the definitional problems associated with this domain is surveyed in Mishra (1990, p. 123 f.).
3. Some authors have drawn up tables in which the dates of origin of highly developed integral social protection systems are presented in chronological order. Averages are calculated in order to facilitate comparisons between the points in time at which individual welfare state elements were introduced. A typical table can be found in Wagschal (2000, p. 39). In this particular instance there are 23 placings for OECD countries. The first five placings are

occupied by Germany, Denmark, Belgium, Austria and Great Britain, while the last five are occupied by Japan, Portugal, the U.S.A., Switzerland and Canada.

4. Cf. Grossekettler (2002, pp. 49-60).

5. This is particularly true of hypotheses which are functionalistic in a methodological sense. A good example is provided by a hypothesis set up by Weisser (1956, p. 398): the so-called "law of parallelism between the processes of industrialization and social security." If we are to believe Weisser, industrialization destroys the protection systems of agricultural societies, thereby "forcing" people to evolve modern social insurance systems. This hypothesis fails to demonstrate the modus operandi and cannot explain why, for instance, the British set up a social insurance system later than the Germans although Great Britain was industrialized much earlier than Germany. Hypotheses about the modus operandi are discussed in e.g., Flora and Alber (1982).

6. For his theory of moral sentiments Smith drew upon a philosophical debate about the nature of human urges. Main contributors to this debate were Hobbes, Hume, Hutcheson, Locke and Mandeville. On the origins and reception of Smith's "Theory of Moral Sentiments" cf. Ross (1995, pp. 157-194).

7. The term is here used in a broad sense.

8. This is borne out by experience.

9. We mean anthropoid apes as well as humans.

10. People who are willing to support others may be motivated by feelings of affection or by self-interest (i.e., the hope of obtaining some kind of service in return).

11. Cf. the literature cited in Grossekettler (2002, p. 54).

12. In order to obviate misunderstandings, a caveat must be added here. Nobody denies that trial and error can be used successfully to discover modes of behavior which bring benefits to *individuals;* nor is there any doubt that such behavior can be systematically imitated. This has been confirmed by research on primates. Thus, for instance, it has been demonstrated that regional differences between chimpanzee "cultures" are sometimes due to serendipitous discoveries made by individual animals (de Waal, 2002). However, when we talk about insurance- and growth-promoting functions, we are concerned with organizational improvements which can only be effected *collectively.* As has already been pointed out, the author has been unable to find any scientific literature on this subject, though some relevant information may have been included in a work entitled *Handbook on the Economics of Giving, Reciprocity and Altruism* (Gerard-Varet et al., to be published in (in prep.)).

13. Similar views are put forward by Nipperdey (1998, p. 339 f.) and Ritter (1989, p. 86 f.).

14. World Bank (1994, p. 315 ff.) distinguishes the same phases with regard to old-age-pension systems. The criteria applied by the Bank include implicit yields and contribution rates which are considered typical of particular phases.

15. As far as the relation of social spending to GDP is concerned, Germany usually occupies one of the first five placings. The following figures for Germany are from Wagschal (2000, p. 40), who compares data from 23 OECD countries, and who centres his attention on the years 1950, 1980 and 1995. The average figures for the countries under consideration are indicated in brackets: 14.8% (8.1%)/25.7% (19.1%)/29.6% (23.7%). If we consider the deviations, we find that in 1950 the figures ranged between 3.5% (Japan) and 14.8% (Germany); in 1980 they ranged between 10% (Japan) and 29.8% (Sweden), and in 1995 they ranged between 14.1% (Japan) and 33.4% (Sweden).

16. During this period, sociopolitical expansion and stagnation phases can only be observed in advanced industrial countries. There were no corresponding phenomena in countries such as Greece, Portugal or Spain, whose social insurance systems were still in their infancy.

17. Although we have not adduced any statistical evidence to support our assertion, we can safely state that in many countries quotas began to rise immediately after the war. There

is, of course, a connection between this phenomenon and state responsibility for those who suffered loss or injury in the war. In the international literature on social security the three post-war decades are often referred to as the "golden area" of social policy (cf. Pearson and Scherer, 1997, p. 6).

18. Snower (2000, p. 39) throws interesting light upon the distinction between the two phases under discussion.

19. In order to explain Wagner's Law, Peacock and Wiseman (1961) developed the so-called displacement effect hypothesis. According to this theory, crises such as world wars help to overcome taxpayer resistance to higher taxes, and politicians can therefore raise the public sector's share of GDP. However, once a crisis has been resolved, people become used to a high level of taxation, and the public sector's share of GDP continues to rise instead of falling to the original level. This phenomenon is known as the displacement effect. Since Peacock/Wiseman's theory was elaborated exclusively on the basis of data concerning public expenditure growth in the first half of the twentieth century, many economists now take the line that the displacement effect hypothesis is inadequate in its original form. From the purely empirical viewpoint, Wagner's Law has undoubtedly proved its worth, but space precludes an extended discussion of Wagner's assertions in the present article. Nonetheless, we might add that in recent years some interesting new theories have been expounded in an attempt to account for the expenditure growth predicted by Wagner. Cf. Gwartney et al. (1998), Holsey and Borcherding (1997), Kleinmann (1985), Tanzi and Schuknecht (1997), Theurl (1990), and the specialist literature cited in these studies.

20. Up to now, economists have concentrated on the growth cycle and the long-term development of the unemployment rate, and in the period under review changes in the age structure of the population were of little importance. In future, however, particular attention will have to be paid to long-term demographic developments because of the consequent financial drain on old-age-pension and health care systems.

21. A similar view is expressed by Zimmermann (1996, p. 5).

22. Politicians who introduce welfare measures win popularity with the electors.

23. People gradually realize that it is the taxpayer who has to defray the cost of a social safety net, and there is no escaping the fact that as the population ages, public pension schemes will have to pay out more than they take in.

24. There may be two reasons for this. First, when a new insurance system has just been set up, the insurer does not normally have to pay out large sums of money. Second, boundary conditions may be particularly favorable.

25. This kind of problem may be caused by an unpropitious demographic development or by falling growth. A slowdown in growth may be triggered by exogenous factors, and the situation may be worsened by an ill-conceived social policy. In some cases, a misguided social policy—among other things—may even be at the root of the trouble.

26. When electors finally wake up to the seriousness of the situation, it may be too late to institute reforms.

27. Pierson (1994) gives a particularly apt description of the way politicians behave during periods of growth and periods of recession. He speaks of "politics of credit claiming" and "politics of blame avoidance."

28. Ritter (1983) provides a detailed comparative study of the early history of social insurance in Germany and Great Britain.

29. A liberal welfare state is free market oriented, a conservative one is cooperation oriented, and a Social Democratic one is administration oriented.

30. There is an earlier OECD publication in which other definitions are used, and these definitions vary in the course of time (OECD, 1985). If we were to bring together the two

statistical series, we would observe sudden rises in social expenditure ratios in the early 1980s. These increases would be somewhere in the region of 10%.

31. Responsibility for the payment of benefits such as sick pay often rests with private employers rather than with the state.

32. European Court of Justice, Case C-55/94 (Gebhard), European Court reports 1995, p. I-04165f., marginal number 37.

33. For further information see Wissenschaftlicher Beirat beim Bundesministerium der Finanzen(2000, pp. 9-14).

34. This is the first of the five principles listed above.

35. They miscalculate their insurance requirements because they tend to underestimate the importance of minor risks and of certain events which might occur in the very distant future.

36. Here we should not forget the problems posed by disablement and industrial accidents.

37. Space precludes a discussion of the special problems associated with health care. See Fuchs (1998) for information about the specialist literature on the subject of health economics.

38. Typical examples of such costs are loss of efficiency and lack of growth.

39. In other words, later generations will derive fewer benefits from the yields.

40. This obligation may apply to certain social categories or to the entire population.

41. The National Health Service is a case in point.

42. This is the price of a man hour from the employer's viewpoint.

43. This is the price of a man hour from the employee's viewpoint.

44. The aim of this program is to help participants achieve self-sufficiency through employment. In order to reduce welfare payments to the jobless, the government tries to remove obstacles that prevent people from joining the labor force. An evaluation study of the program has been conducted by the Wisconsin Legislative Audit Bureau (2001).

45. In a fully funded system workers have to save up to finance their own old-age pensions.

46. In a pay-as-you-go system current earnings are immediately converted into pension payments. As a result, the first generation of pensioners is not obliged to build a stock of capital. In this context, attention must be drawn to a fundamental difference between the fully funded and pay-as-you-go systems. It is a difference that plays a particularly prominent role in pension insurance, although it may also be observed in other types of insurance in which the age structure of insurees is important. The difference in question becomes apparent when we consider three distinct solutions to the problems posed by the fact that it is the elderly who derive the greatest benefits from health insurance schemes. The first solution is a system where all insurees are required to pay age-independent contributions that cover current expenditure (a comprehensive pay-as-you-go system). The second solution is a system where insurees have to pay age-dependent contributions to cover cohort expenditure (an age group pay-as-you-go system). The third solution is a fully funded system, where provisions are made in order to cover increased expenditure for the elderly.

47. Orth (2000, p. 106) offers a survey of empirical studies concerning the influence of pension-funding systems on the accumulation of savings. Many authors fail to detect any significant difference between the two funding systems. Börsch-Supan (2001b, p. 6) believes there is fairly persuasive empirical evidence to back hypotheses about a reduction in savings and a capital stock difference effect in Germany, the Netherlands and the U.S.A. In the specialist literature, reference is often made to the so-called Feldstein controversy. But as Homburg (1988, pp. 61-65) has shown, Feldstein (1974) does not make general comparisons between fully funded schemes and pay-as-you-go systems. Instead, he examines the special characteristics of the American system, where pension payments are set off against pensioners' earnings.

48. Kotlikoff (1998, p. 415 f.) discusses the relevance of various savings hypotheses to comparisons between fully funded schemes and pay-as-you-go systems.

49. This debt burden might be viewed as a kind of implicit public debt. For further information see Kotlikoff (1998, p. 416 f.).

50. Space precludes a detailed discussion of the efficiency gains that can be achieved by means of the fully funded system. For further information see Homburg (1988). The fully funded system would be less prone to misuse by politicians because the pay-as-you-go system—unlike the fully funded system—gives the median voter strong incentives to advocate more generous benefits (Homburg, 1988, pp. 115-131).

51. International surveys can be found in the following publications: OECD (1988), Ferge and Kolberg (1992), Lottes (1993), World Bank (1994), Franco and Munzi (1996), OECD (1996), Schmid (1996), Gruber and Wise (1999), OECD (1998b), Orth (2000), Kaufmann (2001), Börsch-Supan and Miegel (2001) and Siebert (2002). The references have been presented in chronolgical order.

52. Germany's senior citizen ratio (the number of people over sixty divided by the number of gainfully employed people aged between 20 and 59) was approximately 36% in 1995, and it will rise to somewhere in the region of 102% in 2050. This means the situation is much more serious in Germany than in any of the other countries whose data have been compared. In the U.S.A., for instance, the senior citizen ratio is expected to rise by 30% to 53%, while in Chile it will probably rise by 18% to 56% (Börsch-Supan, 2001b, p. 8). A recent comparison of a greater number of countries by Jackson and Howe (2003) ascribes Germany a medium vulnerability to aging problems whereas social security systems in France, Italy and Spain are expected to face the most serious threats from aging.

53. This is income during the initial retirement period, divided by earnings during the final period of employment.

REFERENCES

Adema, W. (1997). "What Do Countries Really Spend on Social Policies? A Comparative Note." *OECD Economic Studies* (28): 153-167.

Alber, J. (1999). "Der Deutsche Sozialstaat im Lichte international vergleichender Daten." In *Sozialberichterstattung und Sozialbeobachtung,* Flora, P. and Noll, H.-H., eds., pp. 41-72. Frankfurt am Main/New York: Campus.

Auerbach, A. J. and Kotlikoff, L. J. (1987). "Life Insurance of the Elderly: Adequacy and Determinants." In *Work, Health, and Income among the Elderly,* Burtless, G., ed., pp. 29-67. Washington, D.C.: Brookings Institution.

Auerbach, A. J., Kotlikoff, L. J., and Leibfritz, W. (eds.) (1999). *Generational Accounting Around the World.* Chicago: University of Chicago Press.

Barr, N. (1992). "Economic Theory and the Welfare State: A Survey and Interpretation." *Journal of Economic Literature* XXX: 741-803.

Barro, R. J. (1974). "Are Government Bonds Net Wealth?" *Journal of Political Economy* 82: 1096-1117.

Beveridge, W. H. (1942). *Social Insurance and Allied Services.* Cmnd 6404, London: HMSO.

Börsch-Supan, A. H. (2001a). "The German Retirement Insurance System." In *Pension Reform in Six Countries. What Can We Learn From Each Other?* Börsch-Supan and Miegel, eds., pp. 13-38. Berlin: Springer.

Börsch-Supan, A. H. (2001b). "Six Countries—and no Pension System Alike." In *Pension Reform in Six Countries. What Can We Learn From Each Other?* Börsch-Supan and Miegel, eds., pp. 1-12. Berlin: Springer.

Börsch-Supan, A. H. and Miegel, M. (2001). *Pension Reform in Six Countries. What Can We Learn From Each Other?* Berlin: Springer.

Briggs, A. (1961). "The welfare state in historical perspective." *Archives européennes de sociologie* 2 (2): 221-259.

Bundesministerium für Arbeit und Sozialordnung (ed.) (1998). *Europaatlas. Soziale Sicherheit im Vergleich.* Bonn.

Caminada, V. and Goudswaard, K. (2001). "International Trends in Income Inequality and Social Policy." *International Tax and Public Finance* 8: 395-415.

Craig, G. A. (1999). *Deutsche Geschichte 1866–1945. Vom Norddeutschen Bund bis zum Ende des Dritten Reiches.* Munich: Beck.

de Waal, F. (2002). "Wir neigen dazu, uns zu überschätzen. Der Affenforscher Franz de Waal über die Kultur der Tiere und die Natur des Menschen." *Der Spiegel* 33: 140f.

Deutscher Bundestag (ed.) (1996). *Fragen an die deutsche Geschichte. Wege zur parlamentarischen Demokratie, Katalog zur historischen Ausstellung im Deutschen Dom in Berlin,* 19 ed. Bonn: Deutscher Bundestag.

Diamond, P. A. (1977). "A Framework for Social Security Analysis." *Journal of Public Economics* 8: 275-298.

Diamond, P. A. (1998). "Issues in Social Insurance." In *Frontiers of Research in Economic Theory,* Jacobs, D. P., Kalai, E., and Kamien, M. I., eds., pp. 142-159. Cambridge, Mass: Cambridge University Press.

Diamond, P. A. (1999). *Issues in Privatizing Social Security.* Cambridge, MA, London: MIT Press.

Dixon, J. (1999). "Comparative Social Security: The Challenge of Evaluation." *Journal of Comparative Policy Analysis: Research and Practice* 1: 61-95.

Eibl-Eibesfeldt, I. (1997). *Die Biologie des menschlichen Verhaltens. Grundriß der Humanethologie,* 3 ed. Weyarn: Seehamer.

Esping-Andersen, G. and Korpi, W. (1984). "Social policy and Class Policies in Post-War Capitalism: Scandinavia, Austria and Germany." In *Order and Conflict in Contemporary Capitalism,* Goldthorpe, J. H., ed. Oxford: Clarendon Press.

European Commission (1996). *Soziale Sicherheit in Europa 1995.* Brussels/Luxembourg.

Eurostat (1996). *ESSPROS Manual.* Luxembourg.

Eurostat (2000). *Social Protection Expenditures and Receipts (Data 1980–1998).* Luxembourg.

Feldstein, M. (1974). "Social Security, Induced Retirement, and Aggregate Capital Accumulation." *Journal of Political Economy* 82: 905-920.

Ferge, Z. and Kolberg, J. E. (eds.) (1992). *Social Policy in a Changing Europe.* Frankfurt a.M.: Campus/Boulder: Westview Press.

Flora, P. and Alber, J. (1982). "Modernization, Democratization, and the Development of Welfare States in Western Europe." In *The Development of Welfare States in Europe and America,* Flora and Heidenheimer, eds., pp. 37-80. New Brunswick: Transaction Books.

Flora, P. and Heidenheimer, A. J. (eds.) (1982). *The Development of Welfare States in Europe and America.* New Brunswick: Transaction Books.

Folkers, C. (1983). *Begrenzungen von Steuern und Staatsaufgaben in den USA.* Baden-Baden: Nomos.

Franco, D. and Munzi, T. (1996). "Public Pension Expenditure Prospects in the European Union: A Survey of National Projections." *European Economy* 3: 1-126.

Frank, R. H. (1992). *Die Strategie der Emotionen.* Munich: Oldenbourg.

Frerich, J. and Frey, M. (1996a). "Handbuch der Geschichte der Sozialpolitik in Deutschland." *Bd. 1: Von der vorindustriellen Zeit bis zum Ende des Dritten Reiches,* 2 ed. Munich/Vienna: Oldenbourg.

Frerich, J. and Frey, M. (1996b). "Handbuch der Geschichte der Sozialpolitik in Deutschland." *Bd. 3: Sozialpolitik in der Bundesrepublik Deutschland bis zur Herstellung der Deutschen Einheit, 2* ed. Munich/Vienna: Oldenbourg.

Fuchs, V. R. (1998). "Health economics." In *The New Palgrave. A Dictionary of Economics,* Eatwell, J., Milgate, M., and Newman, P., eds., Vol. 2, pp. 614-619. London: Macmillan, New York: Stockton Press.

Gerard-Varet, L.-A., Kolm, S. C., and Mercier Ythier, J. (eds.) (in prep.). *Handbook on the Economics of Giving, Reciprocity and Altruism.* Amsterdam: Elsevier.

Gough, I. (1998). "Welfare state." In *The New Palgrave. A Dictionary of Economics,* Eatwell, J., Milgate, M., and Newman, P., eds., Vol. 4, pp. 895-897. London: Macmillan, New York: Stockton Press.

Grossekettler, H. (2002). "Ursprünge und Entwicklungslinien sozialer Sicherungssysteme in Deutschland." In *Arbeitsmärkte und soziale Sicherungssysteme unter Reformdruck,* Apolte, Th. and Vollmer, U., eds., pp. 47-81. Stuttgart: Lucius & Lucius.

Gruber, J. and Wise, D. (1999). *Social Security and Retirement around the World.* Chicago: The University of Chicago Press.

Gwartney, J., Holcome, R., and Lawson, R. (1998). "The Scope of Government and the Wealth of Nations." *Cato Journal* 18: S. 163-190.

Harris, M. (1995). *Kannibalen und Könige. Die Wachstumsgrenzen der Hochkulturen.* Munich: Deutscher Taschenbuch Verlag.

Hennock, P. (1982). "Die Ursprünge der staatlichen Sozialversicherung in Großbritanien und das deutsche Beispiel 1880–1914." In *Die Entstehung des Wohlfahrtsstates in Großbritannien und Deutschland, 1850–1950,* Mommsen, W. J., ed., pp. 92-114. Stuttgart: Klett-Cotta.

Hertner, P. (2002), "Entstehung und Wandel von Arbeitsmarktinstitutionen in Deutschland von der Mitte des 19. Jahrhunderts bis zur Weltwirtschaftskrise 1929/33." In *Arbeitsmärkte und Soziale Sicherungssysteme unter Reformdruck,* Apolte, Th. and Vollmer, U., eds., pp. 27-45. Stuttgart: Lucius & Lucius.

Holsey, Ch. M. and Borcherding, Th. E. (1997). "Why does government's share of national income grow? An assessment of the recent literature on the U.S. experience." In *Perspectives on Public Choice,* Mueller, C., ed., p. S.562-589. Cambridge: Cambridge University Press.

Holzmann, R. and Jørgensen, S. (2001). "Social Risk Management: A New Conceptual Framework for Social Protection, and Beyond." *International Tax and Public Finance* 8: 529-556.

Homburg, St. (1988). *Theorie der Alterssicherung.* Berlin: Springer.

Homburg, St. (1990). "The Efficiency of Unfunded Pension Schemes." *Journal of Institutional and Theoretical Economics* 146: 640-647.

Homburg, St. and Richter, W. F. (1990). "Eine effizienzorientierte Reform der GRV." In *Bevölkerung und Wirtschaft,* Felderer, B., ed., pp. 183-191. Berlin: Duncker & Humblot.

Hopmann, R. (1998). *Eine vergleichende Analyse der Finanzierung von Sozialsystemen unter allokativen Gesichtspunkten.* Frankfurt/Berlin/Bern/New York/Paris/Vienna: Lang.

Jackson, R. and Howe, N. (2003). *The 2003 Aging Vulnerability Index. An Assessment of the Capacity of Twelve Developed Countries to Meet the Aging Challenge.* Washington D.C./London: Center of Strategic and International Studies and Watson Wyatt Worldwide.

Kaufmann, F.-X. (2001). "Der deutsche Sozialstaat im internationalen Vergleich." In *Geschichte der Sozialpolitik in Deutschland seit 1945,* Hockerts, H. G., eds., pp. 799-989. Baden-Baden: Nomos.

Kleinmann, K. (1985). *Die langfristige Entwicklung von Kommunikation und Staatsausgaben.* Frankfurt/Berlin/Bern/New York/Paris/Vienna: Lang.

Klengel, H. (1999). *König Hammurapi und der Alltag Babylons.* Düsseldorf/Zürich: Artemis & Winkler.

Kotlikoff, L. J. (1998). "Social Security." In *The New Palgrave. A Dictionary of Economics,* Eatwell, J., Milgate, M., and Newman, P., eds., Vol. 4, pp. 413-418. London: Macmillan, New York: Stockton Press.

Lampert, H. (1998). *Lehrbuch der Sozialpolitik,* 5 ed. Berlin: Springer.

Lassila, J. and Valkonen, T. (2001). "Pension Prefunding, Aging, and Demographic Uncertainty." *International Tax and Public Finance* 8: 573-593.

Lottes, G. (ed.) (1993). *Soziale Sicherheit in Europa, Renten- und Sozialversicherungssysteme im Vergleich.* Heidelberg: Physica-Verlag.

Mannheim Research Institute for the Economics of Aging (2002). "Artikel 'Rentenreformen— die Meinung der Deutschen.' "*Newsletter* 9/2002: 2-4; http://www.mea.uni-mannheim.de/ mea_neu/pages/files/newsletter/newsletter04.pdf.

Metze, I., Prinz, A., and Lübke, E. (2003). "Alternde Bevölkerung und Rentenreform im Mehrländermodell." In *Agglomeration, Population und Koordination in Europa,* Prinz, A., Steenge, A., and Vogel, A., eds., pp. 133-162. Münster/Hamburg/London: Lit.

Mirrlees, J. A. (1995). "Private Risk and Public Action: The Economics of the Welfare State." *European Economic Review* 39: 383-397.

Mishra, R. (1990). *The Welfare State in Capitalist Society. Policies of Retrenchment and Maintenance in Europe, North America and Australia.* New York: Harvester Wheatsheaf.

MISSOC (2001). Mutual Information System on Social Protection in the EU Member States and the EEA, Social Protection in the Member States in the EU Member States and the European Economic Area. Situation on January 1st 2001 and Evolution, http://europa.eu.int/comm/employment_social/missoc2001/index_en.htm.

Nipperdey, Th. (1998). *Deutsche Geschichte 1866–1918, Vol. 1: Arbeitswelt und Bürgergeist,* Munich: Beck.

OECD (1985). *Social Expenditure 1960–1990. Problems of Growth and Control.* Paris: OECD.

OECD (1988). *Ageing Populations: The Social Policy Implications.* Paris: OECD.

OECD (1996). *Social Expenditure Statistics of OECD Members Countries. Provisional Version.* Paris: OECD.

OECD (1998a). *1980–1998. 20 Years of Social Expenditure, The OECD Database.* http://www. oecd.org/pdf/M00028000/M00028861.pdf.

OECD (1998b). *Family Ressources During Retirements, Aging Working Papers AWP 4.3.* Paris: OECD.

Orth, A. (2000). *Der Einfluß unterschiedlicher Formen von Alterssicherungssystemen auf das Kapitalmarktaufkommen: Eine empirische Analyse der Daten und Institutionen ausgewählter OECD-Staaten,* Frankfurt am Main: Lang.

Ott, N. (2000). "Ökonomische Effizienz und Familienlastenausgleich—eine konzeptionelle Klärung." In *Familienwissenschaftliche und familienpolitische Signale,* Jans, B., Habisch, A., and Stutzer, E., eds., pp. 185-200. Grafschaft: Vektor Verlag.

Peacock, H. and Wiseman, J. (1961). *The Growth of Public Expenditure in the United Kingdom.* Princeton: Princeton University Press.

Pearson, M. and Scherer, P. (1997). "Balancing Security and Sustainability in Social Policy." *The OECD Observer* 205: 6-9.

Pierson, P. (1994). *Dismantling the Welfare State? Reagan, Thatcher, and the Politics of Retrenchment.* Cambridge: Cambridge University Press.

Ridley, M. (1996). *The Origins of Virtue.* London: Viking.

Ritter, G. A. (1983). *Sozialversicherung in Deutschland und England: Entstehung und Grundzüge im Vergleich.* Munich: Oldenbourg.

Ritter, G. A. (1989). *Der Sozialstaat. Entstehung und Entwicklung im internationalen Vergleich, Historische Zeitschrift, Beiheft 11.* Munich: Oldenbourg.

Ritter, G. A. (1997). "Sozialpolitik im Zeitalter Bismarcks. Ein Bericht über neue Quelleneditionen und neue Literatur." *Historische Zeitschrift* 265: 683-720.

Rolf, G., Spahn, P. B., and Wagner, G. (eds.) (1988). *Sozialvertrag und Sicherung: Zur ökonomischen Theorie staatlicher Versicherungs- und Umverteilungssysteme.* Frankfurt/M./New York: Campus, Verlag.

Roscher, W. (1894). *System der Armenpflege und Armenpolitik.* Stuttgart: Cotta.

Rosner, P. G. (2003). *The Economics of Social Policy.* Cheltenham, UK/Northhampton, MA: Edward Elgar.

Ross, I. S. (1995). *The Life of Adam Smith.* Oxford: Clarendon Press.

Rothschild, M. and Stiglitz, J. E. (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics* 90: 630-49.

Saile, W. (1958). *Hermann Wagener und sein Verhältnis zu Bismarck. Ein Beitrag zur Geschichte des konservativen Sozialismus.* Tübingen: Mohr.

Schewe, D. (2000). *Geschichte der sozialen und privaten Versicherung im Mittelalter und den Gilden Europas.* Berlin: Duncker & Humblot.

Schmid, J. (1996). *Wohlfahrtsstaten im Vergleich. Soziale Sicherungssysteme in Europa: Organisation, Finanzierung, Leistungen und Probleme.* Opladen: Leske + Budrich.

Schmidt, M. G. (1988). *Sozialpolitik. Historische Entwicklung und internationaler Vergleich.* Opladen: Leske + Budrich.

Siebert, H. (ed.) (2002). *Economic Policy for Aging Societies.* Berlin: Springer.

Siegel, N. A. (2000). Jenseits der Expansion? Konsolidierung und Rückbau der Sozialpolitik. Theorien und Befunde aus vergleichender Perspektive, Diss. Bremen, http://www.rzuser.uni-heidelberg.de/~ds8/Diss_online.htm.

Sinn, H.-W. (1998). *The Pay-as-you-go Pension System as a Fertility Insurance and Enforcement Device, CES Working Paper No. 154.* Munich: CES.

Sinn, H.-W. (2000). "Sozialstaat im Wandel." In *Die Zukunft des Sozialstaats, Jahrestagung des Vereins für Socialpolitik, 1998,* Hauser, R., ed., pp. 15-34. Berlin: Duncker & Humblot.

Smith, A. (1976/1759). "The Theory of Moral Sentiments." In *The Glasgow Edition of the Works and Correspondence of Adam Smith,* Raphael, D. D. and Macfie, A. L., eds., Vol. 1. Oxford: Clarendon Press 1976, first edition 1759.

Snower, D. J. (2000). "Evolution of the Welfare State." In *Die Zukunft des Sozialstaats, Jahrestagung des Vereins für Socialpolitik, 1998,* Hauser, R., ed., pp. 35-52. Berlin: Duncker & Humblot.

Stein, L. von (1855). *Das Königtum, die Republik und die Souveränität der französischen Gesellschaft seit der Februarrevolution 1848,* Vol. 3 der "Geschichte der sozialen Bewegung in Frankreich von 1789 bis auf unsere Tage," 2 ed. Leipzig: Otto Wigand.

Tanzi, V. and Schuknecht, L. (1997). "Reforming Government: An Overview of Recent Experience." *European Journal of Political Economy* 13: 395-417.

Theurl, E. (1990). "Gesetze der Entwicklung der Staatstätigkeit: Der Transaktionskostenansatz." *WiSt* 19: S.503-507.

Volkert, J. (1999). *Uunter Mitarbeit von E.-M. Schick: Soziale Dienste und Umverteilung in Deutschland.* Berlin: Duncker & Humblot.

Wagschal, U. (2000). "Besonderheiten der gezügelten Sozialstaaten." In *Der gezügelte Wohlfahrtsstaat. Sozialpolitik in reichen Industrienationen,* Obinger, H. and Wagschal, U., eds., pp. 37-72. Frankfurt am Main/New York: Campus.

Weisser, G. (1956). "Soziale Sicherheit." In *Handwörterbuch der Sozialwissenschaften,* Beckerath, E. V. et al., eds., Vol. 9, pp. 396-412. Stuttgart: Fischer, Tübingen: Mohr, Göttingen: Vanderhoeck & Ruprecht.

Wendt, B. J. (1993). "'Sozialstaat' und 'Welfare State'—unterschiedliche Traditionen im Vergleich." In *Soziale Sicherheit in Europa,* Lottes, G., ed., pp. 29-55. Heidelberg: Physica-Verlag.

Wildavsky, A. (1980). *How to Limit Government Spending.* Los Angeles/Berkley: University of California Press.

Wildavsky, A. and Caiden, N. (2001). *The New Politics of the Budgetary Process.* New York/London: Longman.

Wilson, E. O. (1976). *Sociobiology. The New Synthesis.* Cambridge, Mass./London: Harvard University Press.

Wisconsin Legislative Audit Bureau (2001). Wisconsin works (W-2) Programm. An Evaluation, http://www.legis.state.wi.us/lab/reports/01.7tear.htm.

Wissenschaftlicher Beirat beim Bundesministerium der Finanzen (2000). *Freizügigkeit und Soziale Sicherung in Europa.* Bonn: Stollfuß.

World Bank (1994). *Averting the Old Age Crisis: Policies to Protect the Old and Promote Growth.* Oxford: Oxford University Press.

Zimmermann, H. (1996). *Wohlfahrtsstaat zwischen Wachstum und Verteilung.* Munich: Vahlen.

Zimmermann, H. and Henke, K.-D. (2001). *Finanzwissenschaft,* 8 ed. Munich: Vahlen.

Zwiedineck-Südenhorst, O. von (1911). *Sozialpolitik,* Leipzig/Berlin: Teubner.

Chapter 14

REDISTRIBUTION, POOR RELIEF, AND THE WELFARE STATE

Richard E. Wagner
George Mason University
rwagner@gmu.edu

Abstract   This chapter examines various justificatory arguments regarding state involvement in redistribution and poor relief, and contrasts those arguments with some of the implications for redistribution that emerge out of rent seeking models of government. This examination includes both historical consideration and contemporary controversy, and covers territory ranging from the state guarantee of income to full reliance on private provision.

Keywords: Poor relief, redistribution, welfare state, rent seeking

JEL classification: D31, D63, H23, I38

The redistributive programs that constitute the welfare state have become objects of growing controversy. This controversy involves contrasting claims about both the impact of a market economy on poverty and the contribution, positive or negative, of government programs to poverty. Those who support an extensive welfare state typically claim that however strongly a market economy might promote economic progress, it also leaves behind a good number of people in its progressive wake. In sharp contrast, there is a good deal of argument and evidence in support of claims that governments do much to impede progress and promote poverty. To the extent these contrary claims are correct, an effective program of poor relief would seem to require a less energetic welfare state than we now see.

The programs of the welfare state are often described as forming a type of safety net. It is hard to object to a safety net. Among people who are trying to climb high to the best of their talents, some may fall through no fault of their own. If they do, the safety net breaks their fall and sends them on their

way again. This is a vision of the welfare state as offering people a helping hand if needed to support their own responsible conduct, as distinct from giving a handout that substitutes for responsible conduct. Much of the growing controversy over the welfare state arises over whether the hand or the handout is the more accurate vision. Various general examinations of this controversy are presented in Atkinson (1999), Barr (1993), Beito (2000), Ebeling (1995), Mead (1986), Murray (1984), Offe (1984), Olasky (1992), Rector and Lauber (1995), Schmidtz and Goodin (1998), Tanner (1996), Tullock (1983), Wagner (1989), and Weicher (1984).

There is no doubt that people try in many ways to protect themselves against disruptive or calamitous events. The development of insurance is good testimony to the energy and creativity that people have brought to the search for such protection. Through insurance people cover themselves against a variety of catastrophes ranging from accident or illness through the destruction of property. People can also create safety nets through saving, which can support them against unemployment and provide annuities for retirement. To be sure, a safety net is not created through individual effort alone. Everyone is born into a family, and families are valuable sources of support and instruction. So too are churches and a variety of associations and organizations that people create to deal with their needs for mutual support. In any case, people will craft safety nets on their own without government.

The claims on behalf of the welfare state are that the state can supplement and support the other efforts of people in society. The welfare state thus fills in gaps in the safety nets that people create for themselves. A considerable body of analysis and evidence, however, tells a different story. This alternative story is one where the welfare state does not seem so much to complement or support individual effort and initiative as it seems to undermine it. We might assert that dependency or poverty is a function of the size of welfare programs, as illustrated by $D = f(W)$. One possibility is that $f' < 0$, indicating that dependency or poverty varies inversely with welfare spending. Most supporters of an expansive welfare state would advance some form of this claim. Another possibility is that $f' > 0$, which would indicate that dependency varies directly with welfare spending. Most of those who support some contraction of the welfare state advance some form of this claim.

Furthermore, dependency is not the same thing as poverty. Welfare spending might reduce poverty while increasing dependency. Indeed, one of the primary lines of argument against an expansive welfare state is that its programs induce people to rely less on their own efforts and more on the state. John Maynard Keynes (1951), in his biographical essay on William Stanley Jevons, notes that Jevons thought that the natural course of the development of civilization would be to eliminate poverty and poor relief as a source of concern. In particular, Keynes cited an 1869 address that Jevons made to the Manchester Statistical

Society. In that address, Jevons lamented how medical charities "nourish in the poorest classes a contented sense of dependence on the richer classes for those ordinary requirements of life which they ought to be led to provide for themselves (p. 301)." Furthermore, Jevons continued "We cannot be supposed yet to have reached a point at which the public or private charity of one class towards another can be dispensed with, but I do think we ought to look towards such a state of things. True progress will tend to render every class self-reliant and independent (p. 301)."

In Jevons' judgment, progress in the organization of economic life would eliminate poverty and dependence. Much of the contemporary criticism of the welfare state holds, similarly with Jevons, an ultimate desideratum of self-reliance within the framework of a market economy, and claims that the welfare state often operates antagonistically to this aim. Many of the supporters of an expansive welfare state seem to reject self-reliance as a desirable end. They seem to aim their gaze instead on a regime where much wealth will be socialized and available to everyone as basic guarantees simply as rewards for being alive, through some form of guaranteed income.

The welfare state is, of course, simply an abstract noun that we use to designate some subset of state programs and activities. James Buchanan (1975) advances the conceptual distinction between the protective and the productive states. The protective state refers to those activities where the state provides and maintains a framework of good civil order within which people can conduct their economic activities. The protective state is a referee that enforces the rules of property and contract which frame and govern economic relationships among people. The productive state refers to the state not as a referee but as a player within the economic process. With respect to Buchanan's dichotomy, the welfare state would seem to involve both roles, at least judging by the supporting rationalizations. One set of justifications claims that the welfare state represents some of the background framework for a market economy. Another set claims that the welfare state represents forms of state production in response to gaps or failures of ordinary market processes and arrangements.

Regardless of whether the welfare state can be represented as subsets of the protective and productive states or treated as a third conceptual category, there is a good deal of vagueness in defining the boundaries of the welfare state. The scope of the welfare state can be defined in quite narrow fashion to include only programs whose clientele is drawn predominately from the poor. This narrow definition of the welfare state would include only a small fraction of the activities that would be covered under a broad definition of the welfare state. A broad definition would include all state programs where some claim about poor relief enters at all into the justifications that people advance for such programs. These days, a broad definition would probably include the predominant share of state activities. Public education, for instance, is not directly

a program of poor relief, but considerations of poverty figure prominently in justifications for public education. It is the same for social security and health care, among numerous other state-supplied services. The same can be said for state regulation, as arguments about poor relief enter into justifications for numerous types of regulations that have little to do directly with poor relief.

This essay starts by reviewing the kinds of justifications that have been given toward using the instruments of public finance as instruments of poor relief. These justifications treat both the tax and the expenditure sides of state budgets, and their point of departure is that the welfare state can "improve" upon the distributive outcomes of a market economy. One approach to justification proceeds on utilitarian grounds by claiming that some degree of redistribution can increase some aggregate measure of utility, and would locate the welfare state as one component of the protective state. Another approach reasons in terms of contracts and claims about market failure, and would locate the welfare state as one component of the productive state. Regardless of the justification that is advanced in support of the welfare state, the state might lack the competence effectively to accomplish what those justifications envision it as accomplishing. There are two broad sets of reasons why this might be so. One is an absence of knowledge about how truly to accomplish what the rationalizations envision it as accomplishing. The other is a lack of interest in actually doing so, perhaps because the force of political interest pulls the state toward other accomplishments. These considerations of competence lead into an exploration of how chasms might arise between the justifications given for addressing poor relief through fiscal measures and the actual consequences of those fiscal measures. Justifications for welfare state redistribution may be the province of fiscal philosophers, but the actual programs of the welfare state are forged in a crucible dominated by fiscal practitioners, political realists all.

1.   UTILITARIAN JUSTIFICATIONS FOR WELFARE STATE REDISTRIBUTION

The dominant strand of argument that fiscal philosophers have advanced for using the state to equalize income is grounded in claims about the utility that people derive from their income. In this regard, primacy of articulation belongs to F. Y. Edgeworth (1897). Suppose a monarch wanted to collect some stipulated amount of revenue from his subjects, and wanted to do so in a manner that imposed the least aggregate sacrifice of utility on his subjects. Revenues are collected in money, but burden is measured in terms of the lost utility that taxes impose on people. If the marginal utility of income is constant, monetary and utility measures are identical. All distributions of a given tax liability among subjects would involve the same aggregate sacrifice
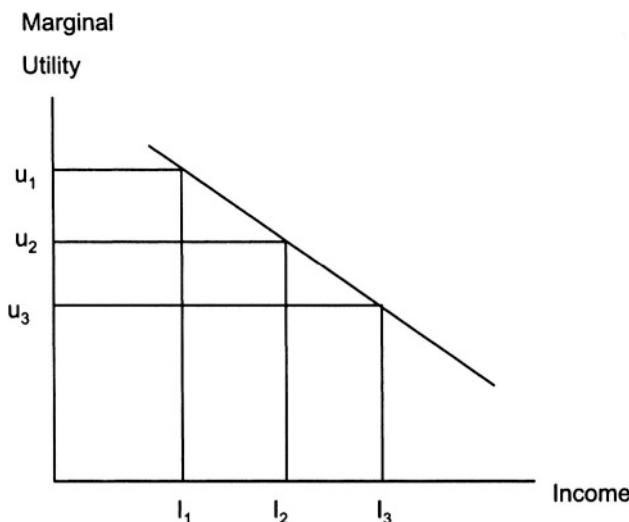
FIGURE 1.

of utility. Most fiscal philosophers, however, have assumed that the marginal utility of income declines with income. This situation is represented by Figure 1, where everyone has the same income-utility schedule but differ in their incomes, and, hence, in the marginal utilities they receive from their incomes. Those incomes are I_1, I_2, and I_3, and the associated marginal utilities are u_1, u_2, and u_3 respectively. A starkly simple conclusion emerges if production or income is independent of the rate of tax. A king who wanted to raise some particular amount of revenue would do so in a manner that pares incomes down from the top. So long as the amount of revenue the king wanted to raise is less than $I_3 - I_2$, he would collect the entire amount from person 3. The amount of equalization that would result would depend on the amount of revenue the king wanted to collect. Full equalization would result once the king's desired revenue reached $[(I_3 - I_1) + (I_2 - I_1)]$.

Rather than minimizing the sacrifices that his revenue demands place upon his subjects, the king's problem could be stated alternatively as one of maximizing the aggregate utility of his subjects. Starting from the position described by Figure 1, a tax on person 3 that was in turn transferred to person 1 would raise aggregate utility by $I_3 - I_1$. So long as the amount of production is invariant to the rate of tax, full equalization would be required for maximization of aggregate utility. Taxes would be imposed on people with above-average incomes, at 100 percent marginal rates, with the revenues transferred to those with below-average incomes. The result would maximize aggregate utility, under the stipulated condition that effort supplied and income generated was independent of the rate of tax and subsidy.

The first-draft conclusion of this utilitarian approach is a full equality of income as an unconstrained optimum. Inequality becomes permissible only through a second-best recognition of constraints that arise because 100 percent marginal rates of tax would destroy incentives to produce. Once this negative effect of taxation on incentive is taken into account, there will be a point beyond which increased redistribution through taxation will depress aggregate utility. For instance, Stern (1976) presents estimates based on various assumptions and simulations where marginal tax rates range from 13 to 93 percent. The analogy in this case is how equally to slice a pie when the size of the pie varies inversely with some measure of equality in the distribution of sizes of the slices.

It might be granted that the marginal utility of income declines, only it could also be asserted that people have different income-utility functions. If so, it is conceivable that a person with low income will nonetheless have a lower marginal utility than someone with high income, because the person with the high income has a higher income-utility schedule. This situation might seem to complicate mightily the king's effort to minimize the sacrifices required by his revenue demands. Yet the king's problem might not be so difficult after all, as Abba Lerner (1944) argued. Suppose the king has no way to match utility schedules with people. The king can minimize his errors by assuming that the same income-utility schedule pertains to everyone. This line of argument allows the conventional analysis of the utilitarian tradeoff to proceed, despite the apparent recognition given to the possibility that people differ in their income-utility functions. This line of argument is grounded in randomnization. It fails in the face of some systematic relation between income and income-utility functions, whereby people with high incomes tend to be those with high income-utility functions.

To be sure, not all tax philosophers have supported the principle of progressive taxation, as illustrated by Walter Blum and Harry Kalven (1953). Nonetheless, the recent literature on optimal taxation, surveyed by Mirrlees (1994), takes off from the earlier sacrifice literature in its use of the income-utility construction. The recent literature on optimal taxation carries forward the utilitarian framework of the sacrifice theorizing, and conceives the government budget as a vehicle for maximizing social utility or welfare. This literature seeks to incorporate into its models a recognition that taxes reduce the amount of effort that people will supply. This reduction of effort puts a limit on the amount of redistribution that the utilitarian calculus would call for, as compared with the full equalization that would be supported if taxation had no effect on the amount of effort people supply. What causes many of those efforts to support only mild progression, is the negative effect upon recipients of transfers of higher marginal tax rates. Indeed, one feature of these models is a zero rate of tax applied at the margin to the highest earner in society.

A presumption that the marginal utility of income declines, and that one more dollar gives less utility to a rich person than to a poor person seems intuitively obvious to many people, so obvious that reservations about the measurability and comparability of utility are readily cast aside. Many have sought to buttress this intuition by resort to arguments about the St. Petersburg Paradox. This paradox is the observation that most people will reject actuarially fair gambles. To be sure, not everyone will do so, and some will accept actuarially unfair gambles, as illustrated by their participation in lotteries. Nonetheless, the St. Petersburg Paradox is widely used to buttress claims about a diminishing marginal utility of income. A person who would be unwilling to bet his entire fortune, double or nothing, on a single coin flip would be characterized as having diminishing marginal utility of income. His expected wealth is the same whether he accepts or rejects the gamble. His failure to gamble, along with a finding that to induce him to gamble the expected value of the gamble must be positive is attributed to a diminishing marginal utility of income. This can be illustrated by Figure 1, where I_2 represents the initial position. A person has a 50:50 chance of moving to I_1 or I_3, the average of which is I_2. In expected value terms, I_2 is equal to a 50:50 gamble between I_1 and I_3. One explanation for why someone might prefer I_2 is diminishing marginal utility of income: the amount of utility lost by moving to I_1 exceeds the gain from moving to I_3.

The St. Petersburg formulation is set in a casino. The explanation as to why someone would reject a fair gamble is that the marginal utility derived from the money won would be less than the marginal utility deducted from the money lost. A casino, however, is not the only setting for choice, and may not be the most suitable one for exploring and illuminating commercial conduct. The income-utility formulation would have us universalize from the particular setting the casino represents. There might be good reason for doing this if the casino were thought to capture some universal quality, as against speaking to some particular setting for choice. The universality of St. Petersburg, however, is dubious. Among other things, it would imply that people would prefer that games end in ties, because the added utility from winning would be less than the decreased utility from losing. With respect to Figure 1, we can denote the income axis as the "psychic" income from playing a game. A tie would yield I_2, a win I_3, and a loss I_1.

People embrace games and surely do not avoid them. And they clearly prefer decisive outcomes to ties. There would seem to be an implicit fiscal sociology built into the utilitarian formulation and its presumption of ubiquitous risk aversion. The most desirable state of affairs is a passive equality in consumption, and what prevents the realization of that equality is the pragmatic recognition that equalization imposes a toll through reduced output. Inequality

is a second-best outcome, countenanced only because of its productive conse-
quences (as noted particularly clearly in John Rawls (1971). An alternative
fiscal sociology would be rooted in activity and not consumption. Games must
have winners as well as losers, and all participants will prefer a shot at I_3 even
though this implies the possibility of I_1, as against settling for I_2, a tie. Success
in any activity is meaningful only when failure is also an option, in commercial
life as well as in athletic contests.

The utilitarian focus on consumption, in contrast to an alternative focus on
activity, leads perhaps almost naturally to a placement of sympathy on those
who have little. It is hard to feel sympathy for people whose pantries are full
in the presence of those whose pantries are empty. But why are some pantries
fuller than others? The utilitarian formulation ignores this question, and in so
doing it distorts the central character of the economic process—the application
of effort to provide opportunities for consumption.

There is no doubt that there are differences among people in their general-
ized productive capacities. People differ naturally in their abilities to fill their
pantries, as a form of act of God, as it were. There is also no doubt that much
of the difference in the condition of various pantries is a matter of choice con-
cerning exertion and foresight. To a considerable extent, those who have fuller
pantries are those who have exerted themselves to this end. They have under-
gone a greater disutility of labor and have postponed consumption more fully
than those with emptier pantries.

Where should the sympathy lie? Take that old childhood story of the three
pigs. The pig who built the brick house had a larger opportunity set than the
pig who built the straw house. Should the sympathy lie with the pig with the
straw house, which might call for a program of taxing pigs who build brick
houses to subsidize pigs who build straw houses? This would be a strange and
destructive pattern of sympathy. The pig who built the brick house exerted
much effort in building that house, he underwent great deprivation. The pig
who built the straw house underwent little deprivation. The case for sympathy
would seem to lie on the side of the pig who bore the deprivation and built the
brick house.

This matter of sympathy is reinforced by considerations of prudence. A
progressive tax policy would tax the builders of brick houses to subsidize the
builders of straw houses. That would reduce the stock of brick houses in soci-
ety and increase the supply of straw houses. Consumption opportunities would
be more equal, and the average level of well being would be lower, assuming
the utility of wolves is not entered into the evaluation calculus.

Normative principles surely should not be socially destructive. The utilitar-
ian principle that derives from a focus on relative consumption opportunities
would seem to induce a pattern of sympathy that is destructive. Sympathy, af-
ter all, properly construed, is not at all synonymous with "feeling sorry for."

Rather, it is synonymous with "wishing to see emulated". It may be fine to feel sorry for, but only if that sympathy is joined with a desire not to see that condition emulated. The utilitarian principle of progressive taxation says that it is morally superior to be a pig who builds a house of straw than to be one who builds a house of bricks.

The utilitarian analysis of progressive taxation construes the central tradeoff in the economic process as one between different items of consumption. Indeed, the Lagrangian multipliers that are found in the problem of maximizing utility subject to a budget constraint are commonly interpreted as a marginal utility of income, in that they show the change in utility deriving from a shift in the budget constraint. In this formulation, some people simply have higher incomes than others. These differences in incomes are the analytical points of departure. The focus is thus on the larger opportunity sets that some people face relative to others, and on the utility associated with different opportunity sets.

Where do endowments come from? In the analysis of consumer choice, from which the utilitarian analysis derives, they are simply there by assumption. This, of course, is impossible. Opportunity sets do not fall from heaven like manna. Opportunity sets must be created through exertion. The fundamental tradeoff in the economic process is not between different, valued items of consumption. Prior to the ability to choose among such items must lie a choice of how much exertion to make and along what directions, so as to make consumption possible.

Crusoe and Friday do not choose between fruit and fish, with one simply facing a larger opportunity set than the other. By doing nothing, they consume nothing, save for dead fish that might wash up on the beach or rotten fruit that might fall to the ground. To advance beyond that rude state of life, exertion is necessary, both directly as in the supply of labor services and indirectly as in the creation of capital goods. The opportunity set starts at the origin. The fundamental economic choice is not between two goods, but between a good and a bad. What is common to both fruit and fish is the exertion that must be undertaken to make consumption possible.

Crusoe and Friday can differ in their consumption possibilities for two types of reasons. One is a natural dominance along all relevant dimensions. Friday might be naturally quicker than Friday at both catching fish and picking fruit. The other is a choice of exertion and providence. Crusoe might devote more time to catching fish and to picking fruit than Friday. He might also devote more effort to creating capital goods than does Friday.

An alternative fiscal sociology would place sympathy on the side of exertion and providence. To be sure, some might question whether success stems from exertion and providence. Among equally situated people, this might be granted. Such people might be thought to have faced similar opportunities for

economic success, only some made better choices than others. But surely peo-ple differ in their initial opportunities. This gets back to the matter of endow-ments. Some people are born with superior opportunities relative to others. Actual income is a mixture of exertions and opportunities. The utilitarian ar-gument treats opportunity as all that matters. The other pole would treat exer-tion and providence as all that matters. Reality undoubtedly lies somewhere in between.

2.   CONTRACTUAL ARGUMENTS FOR WELFARE STATE REDISTRIBUTION

The fiscal literature contains a number of arguments in support of redistri-bution through state expenditure that use a contractual rather than a utilitar-ian analytical framework. These arguments are grounded in claims of market failure, in one form or another. Harold Hochman and James Rodgers (1969) articulated a model that was grounded in the assertion that poor relief had char-acteristics of a public good. In their framework, poor relief would be under-supplied through private charity and related market arrangements. Some state supply would be necessary to secure a Pareto-efficient amount of redistribu-tion, hence they titled their article "Pareto Optimal Redistribution."

Hochman and Rodgers postulated a unidirectional form of utility interde-pendence. Suppose people can be classified as either poor or rich. Hochman and Rodgers postulated that the utility of the poor person depended on his consumption alone, while the utility of the rich person depended both on his consumption and the consumption of the poor person. For the rich person, his utility increased with increases both in his own consumption and in the con-sumption of the poor person. There would be some utility maximizing choice whereby the rich person would make transfers to the poor person until the mar-ginal utility he derives from an increase in the consumption of the poor person equals the marginal utility he derives from his own consumption. Stated in this manner, what exists is simply a choice-theoretic expression of private charity.

An argument for state provision of poor relief enters through the particular presumption advanced about the particular way in which the rich person de-rives utility from poor relief. If the rich person's utility derives simply from the fact of making a transfer, no argument for collective provision emerges. Char-ity would be a purely private good. To convert charity to some form of collec-tive good, it is necessary to postulate that the utility that donors derive from charity depends on the aggregate amount of donations made, or, equivalently, on the aggregate extent to which destitution is reduced. With this alternative formulation in place, poor relief takes on the characteristics of a collective good, and possibilities for free riding emerge. In this formulation, potential individual donors face a form of prisoners' dilemma. Each donor would prefer

to make some contribution to poor relief in conjunction with all other potential donors doing the same. It is individually rational, however, for each donor to withhold his own contribution because it has an imperceptible impact on the aggregate volume of donations. What results is a claim that poor relief is a collective good that will be under-supplied through private charity.

It is a relatively simple matter to advance a claim that the state should act to reduce destitution. There are an indefinitely large number of models that could be constructed to support such a claim (see, for instance, Kliemt (1993) and Wessels (1993), as well as Pasour (1994) to the contrary). To construct such a model does not, of course, make the model correct. The model may have the state acting optimally or efficiently to alleviate destitution, but the state may lack the competence actually to do this. This possible lack of competence has two dimensions: knowledge and incentive.

With respect to knowledge, state officials may not be able truly to determine the efficient amount by which to relieve destitution, or to determine the efficient method or approach. One possible argument against private charity, for instance, is not that it leads to an under-supply of assistance, but that it generates an over-supply of destitution. According to this classical model of poor relief, the nationalization of poor relief reduces destitution, but does so in an entirely different matter from that envisioned in the Paretian approaches to redistribution. The differences in these approaches reflect sharply different claims about the nature of reality as it relates to poverty and poor relief (Himmelfarb, 1983, 1992).

With respect to incentive, state officials may be poorly motivated to alleviate poverty, even if they were secure in their knowledge. It is surely a reasonable presumption that societal processes are dominated by an organized intensity of interest and effort. What gets produced, in the policy arena as elsewhere in society, is dominated by passion and energy, and not by indifference and lassitude. The collectivization of poor relief may fare less well once these considerations of political interest are introduced.

2.1. Knowledge and State Competence

Much of the controversy over poverty and public policy stems from different beliefs about the sources of poverty. Poverty can arise involuntarily, as a matter of *chance,* as representing the luck of the draw or as being an accident of birth. "There but for the grace of God go I" is an expression of this sense. It is surely unreasonable to hold people responsible for their poverty when it arises through one of Nature's involuntary lotteries. Policy prescriptions in such cases would seem almost naturally to run in terms of programs of income redistribution. Indeed, such programs could be construed as a form of social insurance against poor luck, through which the differential bestowal of God's grace is rearranged, as it were.

Alternatively, poverty may result through personal choice (Friedman, 1953). People can choose directly to be poor, as it were, as through foregoing a full-time job to have more time for fishing, or in refusing to attend evening classes three nights a week for six months to qualify for a steady job. But they can also do so indirectly as a by-product of other choices, as in getting pregnant and dropping out of school at 16.

Consider Henry Fawcett's (1871, p. 33) tale of Robinson and Smith, both of whom worked for the same wages and had the same number of dependents. "Robinson is extremely prudent, and does everything in his power to set aside some provision for his old age. By dint of constant thrift he is able … to secure for himself an annuity of 5s. a week. Smith never makes the slightest effort to save, but spends every shilling he can spare at the public-house. When the time comes that he is too old to work he … applies to the parish for maintenance." In Fawcett's continuation of the story, Smith is granted 5s. per week. Robinson points out the manifest unfairness of this grant, and asks for a 2s. supplement, which is denied.

To be sure, the distinction between involuntary poverty by chance and voluntary poverty through choice is simpler to make conceptually than it is to implement empirically. Poverty is generally a mixture of choice and chance, and with that mixture varying from case to case. Chance is ubiquitous in all of our lives, starting with the family situations into which we are born. Those born into loving, nurturing homes will get a better start in life than those born into indifferent or malevolent homes.

What would constitute a successful public policy toward poverty? It is often claimed that measures of poverty based on the money earnings of people exaggerate the amount of poverty because those people also have available a large number of programs that award them in-kind benefits that have monetary value. By some measures the incidence of poverty falls roughly in half, once the implicit income offered by such programs is taken into account. It would seem to follow that the only thing preventing the complete eradication of poverty is sufficient government spending. Yet the permanent existence of people living on government support would hardly seem to indicate the elimination of poverty.

More reasonably, poverty would be defined in terms of the ability of people to be self-supporting. And it is here that problems of poverty policy become especially difficult. It might seem reasonable that policy should seek to aid cases were poverty arises out of pure chance, while refraining from aiding cases where poverty arises through choice. The trouble with this prescription is that it cannot be implemented without knowledge of souls and minds. Nature does not generate birthmarks or other signals that allows such categorization. Mistakes will infect any assistance program, even in a world governed by the best of intentions. The receipt of aid by those who are poor through choice will

encourage more such choices. But to withhold aid to prevent such outcomes will imperil those who are poor through chance.

This tragic dimension is present in Fawcett's tale of Robinson and Smith. When Smith reaches retirement age, it seems cruel to deny him some support. After all, Robinson has an annuity and Smith has none. Some redistribution might seem only fair. Yet Smith's poverty was voluntary, the outcome of earlier choices he made. Is it heartless to refuse aid because Smith's poverty is voluntary? Smith might claim, poignantly and truthfully, that he would not have allowed this to have happened to him had he realized the consequences. Should a second chance, so to speak, be given in this case? What would be the point of refusing to aid? It might punish Smith, but what has been done cannot be undone. Might not some show of compassion toward Smith be in order?

A problem in giving an affirmative answer to this question lies in the lessons that are thereby communicated throughout the society when the aid to Smith becomes generalized as a policy principle, as an illustration of what James Buchanan (1977) calls the Samaritan's Dilemma. For the primary lesson then becomes that a failure to provide for the future will not be a burden to be borne by those who so fail, but will partly be shifted onto those who do not. Failure becomes rewarded, success penalized. Giving aid to those who make impoverishing choices will encourage others to do the same, thereby worsening the problem. Yet there is no unmistakable way of separating choice from chance.

The odds of successful separation can perhaps be improved, however, by replacing public with private forms of assistance. Public assistance must be impersonal and bureaucratic, for requirements of fair treatment must be expressible through objectified rules and procedures. Such an approach is not suitable for making discriminating judgments about who genuinely would use a helping hand profitably and who is simply looking for a handout. Privately organized assistance, where those who supply the assistance not only have greater knowledge of local circumstances and the people with whom they are dealing, but also are free to use the tacit knowledge they have but which cannot be reduced to a table in a memo, perhaps offers a better though still far from Utopian option.

2.2.   Incentive and State Competence

A welfare state creates a specific pattern or network of advantages and disadvantages that get translated into supporting interest groups. One obvious point is that there is a welfare bureaucracy, along with supporting private organizations, for which larger budgets are generally preferable to smaller budgets. To be sure, this general setting characterizes the private sector as well. Dentists want people to be more concerned about their teeth and gums, which in turn translates into more business for them. However, dentists have to attract

business in a setting where customers can choose freely to spend their money elsewhere.

Unlike dentists, or anyone else in the private sector, public sector agencies do not have customers in the traditional sense. To the extent you can speak of customers for such agencies, it would be with reference to the legislative committees that oversee those agencies and make recommendations concerning their budgets. In the private sector there are a variety of employment agencies and mental health councellors who provide services that are similar to some of those that are provided within the poverty subset of the public sector. In the private sector, however, the individuals who pay are the customers, and the suppliers of those services must convince, and repeatedly, the customers that their services are more valuable than alternative uses of their money. A health councellor who provided no remedy but sought simply to corral the largest clientele possible might manage to do so, because perfection exists nowhere.

There are, however, systematic reasons why such conduct would have stronger survival value within the framework of a welfare state. In place of the direct competition for consumer dollars, where every consumer is potentially a marginally relevant consumer, there is a political process of budgetary appropriation. Within a private property setting, what is not spent is returned to owners. But in the institutional setting of a welfare state, such residual claimacy is absent. The public sector counterpart of the councellor faces a legislative committee whose members are generally relatively high demanders of the services being provided. The lack of residual claimacy will lead to less effectiveness in the delivery of services, which implies lower rates of remedy than would result within a regime of private property and market competition.

A welfare state creates at least two sets of interest groups that have interests that support the maintenance of poverty and dependence. One set is the provider of services, only with those providers receiving their funds not directly from customers who are free to use their funds elsewhere, but from legislative committees whose members generally are self-selected for a particularly strong interest in the activities they oversee. The other set is the recipients of the services, as noted particularly crisply by Gordon Tullock (1981), who when faced with an option of continued support or elimination of that support will choose continuation and will support the politicians who promise that continuation.

3.   INTEREST GROUPS AND HORIZONTAL REDISTRIBUTION

The economic literature on income redistribution, whether written from a positive or a normative orientation, largely reflects a presumption that income

transfers are uniformly distributed among the members of any particular income class. Normative literature asks how much income should be transferred from people in the upper income classes to people in the lower income classes. Positive literature asks how much redistribution actually occurs, often arguing that much less is actually accomplished than some of the normative arguments might seem to favor. Both types of literature are cast in terms of a redistributive process that is nondiscriminatory among the members of any particular income class. All members of a particular recipient class are presumed to share equally in the gain, just as all members of a losing class are presumed to share equally in the loss.

Most thinking about redistribution runs in terms of transfers from top to bottom, with people arguing about whether a little or a lot of such redistribution occurs. An alternative formulation is "Director's Law," which was articulated by George Stigler (1970). The Director-Stigler formulation portrays income redistribution as flowing from both the upper and lower classes to the middle class. Whatever the direction of redistribution, however, these formulations treat redistribution as a process that is non-discriminatory among the members of any particular income category. If the highest quintile loses, that loss is shared generally by the members of that quintile. If the lowest quintile gains, that gain is shared generally by the members of that quintile. All of these formulations approach redistribution as a transfer from one horizontal class of people to another, and differ only in terms of how much is transferred.

Despite the generally favorable reception that Richard Musgrave's (1959) conceptualization of the distributive branch of government has generally received among scholars, there is no such thing as a "redistribution program" or policy. There is no collective choice of a single, unified program that represents an effort to impose burdens uniformly on the members of some income class, with the proceeds distributed uniformly among the members of some alternative income class. Rather there are numerous particular programs and policies, which may be aggregated after the fact. But each of those programs transfers income among particular subsets of people. Moreover, the people who are members of any particular income category differ in a wide range of respects, including which part of the country they live in, whether they are self-employed or work for someone else, whether or not they have children, their age, the industry in which they work, and so on. Rather than there being some systemic or global articulation of some distributional objective, there are a variety of competing interest groups, some of which will be successful in the effort to become net recipients of transfers and some of which will not—and so will become net donors instead.

The interest group theory of government (surveyed in Tollison, 1988) claims that political programs transfer income from the unorganized many to the organized few. It is conceivable to aggregate across programs to construct

some global estimate of redistribution. The process that produces redistribution, however, is starkly different from what is commonly envisioned. Redistribution emerges out of competition among political coalitions, and broad income categories provide only a weak basis for the formation of coalitions.

Horizontal coalitions conform to much normative exhortation, but such exhortation is not directly relevant to any positive analysis of redistribution. The formation of public policies that influence the distribution of income emerges through a decentralized process of interest group competition, in which vertical coalitions of demanders of legislation gain at the expense of alternative vertical coalitions who constitute the suppliers of legislation.

If the political process of interest group competition generates a set of programs that transfer income from losing groups to winning groups, the aggregate redistribution that results can be understood only in terms of the underlying process that produced that outcome. For instance, if the domestic automobile industry is a winner in the process of interest group competition, the demand for domestic automobiles will increase by virtue of the higher price imposed on competing, imported automobiles. The distribution of the resulting rent will depend on relative supply elasticities, of course, but in any case there will be some process of vertical distribution. Executives of domestic automobile companies may gain, as might shareholders, who in turn might include union pension funds. The suppliers of specialized labor inputs would gain as well, as would specialized suppliers of inputs to the industry.

In the same way, the redistributive losses in this process of interest group competition would also be apportioned in a vertical and not a horizontal fashion. For instance, suppose the textile industry were to be a loser in this process. Industry rents would fall, or would be negative. This loss would be distributed throughout the structure of complementary inputs within the industry. Executives of the firms in the industry would lose, as would shareholders and specialized labor inputs. A pattern of losses would be spread vertically throughout the range of incomes represented within the industry.

Suppose the outcome of the political process is simply an aggregation over a whole set of interest group measures. Each of those measures contains a vertical pattern of winners and losers. This is illustrated in Table 1 for a 15-person, five quintile model. What is shown there is one particular interest group outcome that transfers income from members of group B to members of group A, leaving the five members of group C unaffected. If this transfer program is aggregated by income category, it appears to be a transfer from the middle income classes to the highest and lowest income classes, the opposite of Director's Law. Yet such an aggregative statement totally misconstrues the essential nature of the program, which is a transfer from everyone in group B to everyone in group A.

TABLE 1.
Redistribution Illustrated and Disaggregated, Anti-Director

Quintile	Group A	Group B	Group C	Aggregate
1	+70	−50	0	+20
2	+30	−40	0	−10
3	+30	−40	0	−10
4	+30	−40	0	−10
5	+40	−30	0	+10

This conceptualization of redistribution as being essentially horizontal and not vertical is certainly consistent with what we know about revolution and insurrection, as explained by Gordon Tullock (1974). Revolutions are not about the masses rebelling against the upper classes. They are about everyone in group A winning at the expense of everyone in group B, along with a differential distribution of the gains and losses among the members of the two groups. An army needs both privates and generals. This is no less true for revolutionary groups who are seeking to take power than it is for those who are seeking to stay in power. Similarly, some of the most intense conflict over the coming of industrialization was surely that between landowners and peasants on the one hand and industrial entrepreneurs and urban workers on the other hand. Relatedly, legislatures are inhabited by people from the far upper tail of the income distribution, regardless of party or ideology.

As further illustration, consider a tax bill that both repeals a tax credit for reforestation and provides for transition rules that allows the steel industry to get a refund for unused tax credits that otherwise would have been rendered worthless by repeal of the investment tax credit. The repeal of the credit for reforestation will have a negative impact on many people throughout the timber industry. Likewise, the transition rules will exert a favorable impact throughout the steel industry. An aggregation of these impacts by income categories may well show some particular "pattern," but doing so would misconstrue totally the nature of the process under examination.

Table 2 illustrates the same central point, only it does so in a way that in aggregative terms is consistent with Director's Law that the middle classes gain at the expense of the upper and lower classes. Yet the essential nature of the process is the same as before: group A takes 200 from group B. The only difference between the two situations is that the distribution of the gains and losses within the two affected groups differs from the preceding case. When expressed in terms of aggregates, Table 2 would seem to describe a very different situation from Table 1, and yet there is really no essential difference between the underlying situations portrayed in the two Tables.

TABLE 2.
Redistribution Illustrated and Disaggregated, Pro-Director

Quintile	Group A	Group B	Group C	Aggregate
1	+20	−50	0	−30
2	+50	−40	0	+10
3	+60	−40	0	+20
4	+50	−40	0	+10
5	+20	−30	0	−10

What holds for the preceding illustration surely holds in general within an interest group model of government. In this more general model we have thousands of interest groups, along with thousands of measures that distribute gains and loses among the members of the various groups. It is always possible to aggregate over all these measures and derive some measure of the amount of redistribution that results, as expressed in terms of income classes. But such an expression of the resulting redistribution both falsifies the nature of the political process that produced the observed outcome and neglects the redistributions among the members of any particular income class.

The treatment of income redistribution in economics is misdirected. Both normative and positive literature speaks as if there were a unified program of nondiscriminatory transfers among broad income classes. It is as if there were a single transfer program of nondiscriminatory taxes imposed on the losers and nondiscriminatory subsidies granted to the winners—and with winners and losers defined in terms of membership in some income class. The considerable controversy over income redistribution in both normative controversy over desirable redistribution and positive controversy over the actual extent of redistribution, has almost universally proceeded in terms of this presumption of a nondiscriminatory process of horizontal redistribution.

The organization of an interest group and the sponsorship of legislation that would aid it is an activity that calls for scarce talent and not common labor. Not being labor that is in common supply, such entrepreneurial talent would customarily be associated with people in the upper ranges of the income distribution. Instead of trying to get protection for the domestic automobile industry, domestic automobile executives could lobby for golden parachutes for displaced auto executives. This might concentrate the rents wholly on themselves, but at the cost of reducing strongly the chance of enactment. By sponsoring broader based legislation that confers benefits throughout the industry, support for the legislation is strengthened. Although factory workers might not be able to organize a coalition and lobby as effectively as the executives, they are more numerous and will be included as beneficiaries within the automobile

interest group. While an army must have generals and colonels, it must also have privates and corporals.

If there were a single program of redistribution, nondiscrimination might be a plausible presumption, though even this is not certain. However, there is no single program, as would be represented by the idea of a redistributive budget. Whatever redistribution that results is the result of aggregating thousands of programs, each of which is intensely discriminatory when compared with any standard of horizontally based redistribution. Moreover, in a world of vertical redistribution, any comparison of actual redistribution by income categories would seem to lose all normative significance because the actual outcomes cannot be reasonably related to the normative categories. An aggregate measure that finds some net redistribution in favor of the lower income classes, for instance, will contain many net losers among members of the lower income categories, as well as containing many net gainers among the upper income categories.

In short, the common approach to income redistribution is predicated upon a presumption that governmental outcomes are the product of some single-minded despot who, some would say, is only imperfectly benevolent. Yet the entire congeries of concepts and categories that has come to exist clashes sharply with the central core of the interest group theory of government. To be sure, that theory emphasizes government as a redistributive process. But that process is animated by the interests of well-organized groups and not by some fiscal philosopher's vision of benevolence.

REFERENCES

Atkinson, A. B. (1999). *The Economic Consequences of Rolling Back the Welfare State.* Cambridge, MA: MIT Press.

Barr, N. (1993). *The Economics of the Welfare State,* 2nd ed. Stanford, CA: Stanford University Press.

Beito, D. T. (2000). *From Mutual Aid to the Welfare State: Fraternal societies and Social Services, 1890–1967.* Chapel Hill: University of North Carolina Press.

Blum, W. J. and Kalven, H., Jr. (1953). *The Uneasy Case for Progressive Taxation.* Chicago: University of Chicago Press.

Breyer, F. (1994). "The Political Economy of Intergenerational Redistribution." *European Journal of Political Economy* 10: 61-84.

Buchanan, J. M. (1975). *The Limits of Liberty.* Chicago: University of Chicago Press.

Buchanan, J. M. (1977). "The Samaritan's Dilemma." In *Altruism, Morality, and Economic Theory,* Phelps Edmund, ed., pp. 71-85. New York: Russell Sage.

Colburn, C. B. and Horowitz, J. B. (1998). "Efficiency Costs and the Demand for Income Redistribution," *Public Finance Review* 26: 214-230.

Ebeling, R. (ed.) (1995). *American Perestroika: The Demise of the Welfare State.* Hillsdale, MI: Hillsdale College Press.

Edgeworth, F. Y. (1897; 1958). "The Pure Theory of Taxation." In *Classics in the Theory of Public Finance,* Musgrave, R. A. and Peacock, A. T., eds., pp. 119-136. London: Macmillan.

Fawcett, H. (1871). *Pauperism: Its Causes and Remedies.* London: Macmillan.

Friedman, M. (1953). "Choice, Chance, and the Personal Distribution of Income." *Journal of Political Economy* 61: 277-290.

Himmelfarb, G. (1992). *Poverty and Compassion: The Moral Imagination of the Late Victorians.* New York: Vintage.

Himmelfarb, G. (1983). *The Idea of Poverty.* New York: Alfred A. Knopf.

Hochman, H. M. and Rodgers, J. D. (1969). "Pareto Optimal Redistribution." *American Economic Review* 59: 542-557.

Keynes, J. M. (1951). *Essays in Biography.* New York: W. W. Norton.

Kidd, A. (1999). *State, Society, and the Poor in Nineteenth-Century England.* London: Macmillan.

Kliemt, H. (1993). "On Justifying a Minimum Welfare State." *Constitutional Political Economy* 4: 159-172.

Lerner, A. P. (1944). *The Economics of Control.* New York: Macmillan.

Mead, L. (1986). *Beyond Entitlement: The Social Obligations of Citizenship.* New York: Free Press.

Mirrlees, J. A. (1994). "Optimal Taxation and Government Finance." In *Modern Public Finance,* Quigley, J. M. and Smolensky, E., eds., pp. 213-281. Cambridge: Harvard University Press.

Murray, C. (1984). *Losing Ground: American Social Policy 1950–1980.* New York: Basic Books.

Musgrave, R. A. (1959). *The Theory of Public Finance.* New York: McGraw-Hill, 1959.

Nozick, R. (1974). *Anarchy, State, and Utopia.* New York: Basic Books.

Offe, C. (1984). *The Contradictions of the Welfare State.* London: Hutchison.

Olasky, M. (1992). *The Tragedy of American Compassion.* Washington: Regnery Gateway.

Pasour, E. C. (1994). "Redistribution and Constitutional Political Economy." *Constitutional Political Economy* 5: 81-98.

Rawls, J. (1971). *A Theory of Justice.* Cambridge: Harvard University Press.

Rector, R. and Lauber, W. F. (1995). *America's Failed $5.4 Trillion War on Poverty.* Washington: Heritage Foundation.

Roberts, R. D. (1984). "A Positive Model of Private Charity and Public Transfers." *Journal of Political Economy* 92: 136-148.

Schmidtz, D. and Goodin, R. E. (1998). *Social Welfare and Individual Responsibility.* Cambridge: Cambridge University Press.

Stern, N. H. (1976). "On the Specification of Models of Optimum Income Taxation." *Journal of Public Economics* 6: 123-162.

Stigler, G. J. (1970). "Director's Law of Public income Redistribution." *Journal of Law and Economics* 13: 1-10.

Tanner, M. (1996). *The End of Welfare: Fighting Poverty in the Civil Society.* Washington: Cato Institute.

Tollison, R. D. (1988). "Public Choice and Legislation." *Virginia Law Review* 74: 339-371.

Tullock, G. (1983). *Economics of Income Redistribution.* Boston: Kluwer Nijhoff.

Tullock, G. (1974). *The Social Dilemma: The Economics of War and Revolution.* Blacksburg, VA: Center for Study of Public Choice.

Tullock, G. (1981). "The Rhetoric and Reality of Redistribution." *Southern Economic Journal* 47: 895-907.

Usher, D. (1992). *The Welfare Economics of Markets, Voting and Predation.* Ann Arbor: University of Michigan Press.

Wagner, R. E. (1986). "Wealth Transfers in a Rent-Seeking Polity." *Cato Journal* 6: 155-171.

Wagner, R. E. (1989). *To Promote the General Welfare.* San Francisco: Pacific Research Institute.

Weicher, J. C. (ed.) (1984). *Maintaining the Safety Net.* Washington: American Enterprise Institute.

Wessels, J. H. (1993). "Redistribution from a Constitutional Perspective." *Constitutional Political Economy* 4: 425-448.

Williams, M. and Reuten, G. (1994). "The Political Economy of Welfare and Economic Policy." *European Journal of Political Economy* 10: 253-278.

Zou, H. (1994). "The Spirit of Capitalism and Long-Run Growth." *European Journal of Political Economy* 10: 279-293.

Chapter 15

ECONOMIC ANALYSIS AND EFFICIENCY IN PUBLIC EXPENDITURE

A. Allan Schmid
Michigan State University
schmid@msu.edu

Abstract   This chapter explores property-compatible and state transfers as alternative approaches to social insurance. It covers both the rationales advanced in support of the different approaches and the actual operation of those programs.

Cost-benefit analysis (CBA) is a set of economic tools used to evaluate the budget size and content of public spending and regulation. Economists are divided over the extent to which CBA can be an independent and authoritative guide to political choice. It is widely used in practice by many governments to evaluate large-scale public infrastructure projects such as roads, bridges, airports, harbors, and water control. It is also used for education and health projects. Increasingly it is used to evaluate regulatory actions (see Ch. 9 this volume). Politicians are also divided as to the use of CBA. They are pleased if the analysis supports what is politically attractive and they tend to ignore the analysis if it is not. It is used much more for identifying a list of projects with positive net returns than for ranking projects when budgets are constrained. This review will put cost-benefit analysis tools within a constitutional and institutional framework.[1]

CBA distinguishes economic from financial analysis. Since a major rationale for government expenditure is to improve upon the performance of the private commercial sector, certain things are accounted for in an economic CBA that would not be in a private financial analysis. There are several contenders for how this is to be done. It will be seen that neither a financial nor

an economic analysis escapes judgments of a political character as to what constitutes an improvement in social welfare.

One school of thought promised to make an independent assessment of efficiency without any instruction from the political authority, who was assumed to be interested in maximizing economic product. There have been several attacks against this position. One was an objection on distributive grounds (Ball, 1979), (Campen, 1986), (Adler and Posner, 1999). The hard line defense was to say that distribution was a separate political matter, and for optimal results, government should redistribute income on a lump-sum basis if it so desires, and not interfere with provision of goods and services. If this line could not be held, some economists were willing to have the benefits to some designated parties politically weighted.

I. M. D. Little and J. A. Mirrlees (1974) argued that the analyst could not independently substitute shadow prices for nominal prices affected by taxes, tariffs and exchange controls without asking for the intent of politicians. If the intent of these policies were corrective of income distribution, then it would be value presumptive for the analyst to replace nominal prices. Earlier Little (1957) critiqued the use of consumer surplus in making welfare comparisons. The theoretical tool which Mishan (1988) and others called the distinguishing feature of policy oriented CBA, Little called a context for political decision. This will be elaborated below in a section on non-marginal projects.

The second attack on analysis made independently of political input is relatively new and comes from those who prize environmental goods and services that do not usually have market prices and accrue to future generations. These products (both project inputs and outputs) did not get fully accounted for (Daly, 1991) (Pearce and Nash, 1981). This seemed to provide an opening for political pricing, but the gap was putatively closed by travel cost and hedonic methods and contingent valuation surveys. The faith in authoritative welfare economics seemed to be renewed by the promise that economists could find new ways in which people inadvertently revealed their preferences, or these could be measured and aggregated directly and without the bias that politicians introduced when listening to voters.

Another attack upon independent authoritative analysis came with the work of Robert Sugden and Alan Williams (1978). They argued for a "decision-maker's approach" which required some explicit input from the politicians. Also see (Stilwell, 1999). The argument was strongest with respect to the discount rate, which Sugden and Williams said could only be a matter of political choice and not a matter of data to be discovered by the analyst. This decision-making approach was sharply critiqued by E. J. Mishan (1982) who in a review said that much would be lost if the economist could no longer be regarded as independent and authoritative. This critique was made in spite of the fact that earlier Mishan had argued that for a project to go forward it must

meet both tests of willingness to pay and willingness to sell (Mishan, 1981, p. 163). This advice makes ranking ambiguous without political input. In the 1988 edition of his book, Mishan says, "I virtually forswear earlier endeavours to base the Pareto criterion of economic efficiency on a consensus, or 'virtual constitution', which, if tenable, would render economic calculation, and economic ranking of alternatives, independent of the outcome of the (democratic) political process (Mishan, 1988),"

There has been a revolution in the theoretical welfare literature in the last 20 years that has not yet been fully felt in applied CBA. One its major ideas is the theory of second best. It says that if the prior income distribution is not politically acceptable, no welfare implications can be drawn from present prices. If costless lump-sum redistributions are not available, then redistribution via projects can not be ruled out as inefficient. The same holds for the prices produced by imperfect capital and labor markets in disequilibrium. The gap between theory and application is nowhere better illustrated than in the work of Boadway and Bruce (1984). They demonstrate the limited applicability of first-best theory in a second-best world with many households with diverse preferences. Yet, in their last chapter devoted to CBA they limit their analysis to "projects which have no perceptible effects on the market prices for goods and factors of production in the economy" and assume that "the economy can be treated as if all persons are identical so that no distributive weights are needed (p. 292)." In other words, they apply CBA to a make-believe, first-best world.

The basis for persistent disequilibria in capital and labor markets is another closely related theoretical development. Stiglitz (1987a) argues that information costs imply that markets will be in disequilibrium even if pure competition and no institutional constraints are present. This reasoning suggests that full utilization of resources is not simply a matter of enforcing competitive markets. Government projects and regulations are part of the everyday management of the economy and not created just to fill the breach of an occasional externality or business cycle slump. For example, disequilibrium means that everyone will not have the same marginal rate of time preference, thus necessitating some political resolution of the conflicting preferences.

Giandomenico Majone (1989, p. 15) has labeled authoritarian policy analysis as "decisionism." He says it assumes a unitary decision maker and is not applicable when there are two or more actors with different objectives. Decisionism assumes all conflicts have been settled and choice of projects or regulations is a technical puzzle to be solved rather than a political judgment to be made. Modern political science is often critical of the Wilsonian faith in technocratic government.

A nascent theory of behavioral economics also has implications for the role of technician and politician. The behavioral sciences remind us that people

and their perceptions differ. The analyst is just another observer with her own cognitions and it can't be assumed that conflicts in perception can be solved outside of the political process.

The literature now contains an applied CBA consistent with second-best welfare theory and the reality of a second-best world. The decision making approach has been extended to a fuller political economy approach where political input has a place appropriate for a democratic society and where no self appointed analyst elite usurps representative multi-level government unwittingly or otherwise.

The outline for such an approach can be sketched by considering in turn the major steps in CBA including establishment of a nominal accounting framework, estimation of the production function, direct benefit estimation, evaluations of non-marginal projects, opportunity cost adjustments for imperfect labor markets, time preference in the context of imperfect capital markets, and preferences for uncertainty adjustments. At each step the iterative interaction between political authorities and analyst will be indicated.

1.   INPUT AND OUTPUT CATEGORIES

The first place where judgment is needed for analysis is in choice of the nominal input and output categories. Nothing can be priced unless the physical quantity of a set of characteristics is understood. This taxonomic choice involves the level of detail and aggregation of product characteristics. When is one product different enough to be given a separate name? Cost-effectiveness analysis presumes that the outputs are desirable and just asks how they can be most cheaply produced. But, even elementary cost-effectiveness analysis can not be done without agreement on whether the cost data is comparing products of equivalent quality. Since people differ on how similar is close enough, some political input is needed. Usually the content of product qualities is chosen by reference to statements of objectives in authorizing legislation. Some dialogue is needed between analyst and politician to establish construct validity (the relationship between qualitatively stated objectives and quantitative results). These are matters of judgment and not simply matters for a logician (Majone, 1989, p. 47).

The issue here is analogous to that in industrial organization where one sign of non-competitive behavior is undue product differentiation. But, what is "undue" for one person is not for another and thus legislative and judicial guidance is needed antecedent to cost comparisons. This issue is the same as that involved in choice of program budget categories that facilitate or hinder comparisons between government agencies.

The issues in selecting output categories may be illustrated with the case of education. Most budget information presented to legislators is in terms of

the costs of inputs—books, teachers, buildings and equipment. However, CBA requires information on outputs. What does education produce? The first measures that come to mind might be number of students taught in various subjects, contact and credit hours, and number of graduates. These are some of the measurable outputs but they tell us little about what the students can do. This might be measured by some aptitude tests that measure skill levels. Different tests emphasize different skills. In any case, it can be asked, what these skills are for. The answer might be conceptualized in terms of employment and income of graduates. But, keep in mind that not all of the sought after impacts of public projects result in a change in income. If an impact such as informed citizenship is omitted from an educational project's performance, it is as if the effect were priced at zero. The further we move along the continuum from inputs to intermediate outputs to final impacts on people's lives, there are more substitutes for any one agency's projects. There may be only one agency producing credit hours in music instruction, but there are more that produce music appreciation and performance.

Just as CBA must account for non-marketed outputs, so must it account for non-marketed inputs. And this too requires choice of categories. For example, a flood control project requires as an input certain wetlands. To price this input requires agreement on qualitative categories which might stop only at area of wetlands or of wetlands of different qualities, perhaps differentiated by the plant species involved and depth and frequency of water. Some environmentalists will want one taxonomy and some another. Before one can ask willingness to pay, one must specify the product.

If the input or output of a proposed project is similar to a previously evaluated project, the issue is what features make them similar so that the previous value can be applied to the goods of the new project (external validity). When you compare goods and say they are similar in all relevant dimensions, you are making a judgment about what is relevant, i.e., what makes them valuable. The names and categories of goods call attention to their valued features. In some cases the name itself differentiates goods. For example, in the private sector, different brands of aspirin sell for different prices. Chemically they are the same, but all consumers do not see it that way. An example in public regulation is the wide range in cost-effectiveness in rules that save lives. We save a life in one area and fail to save a life in another that would cost less. Either governments are woefully inconsistent, different lives saved have different values, or different lives are incommensurate. It is clear, for example, that a life saved in a dramatic fearful context (say cancer or a catastrophe) is regarded as more valuable than a life saved in what is regarded as background risk (such auto accidents). Since goods cannot label themselves, their taxonomy has policy implications.

2.   ESTIMATING THE PRODUCTION FUNCTION

There is much technical expertise in experimental design to establish whether the project input caused a change in output (internal validity). But, better designs cost more and some judgment is needed on whether the reduction in threats to internal validity are worth the cost. Definitive randomly assigned treatments (projects) are rare, which necessitates some qualitative judgment on the weight of the evidence behind alternative projects. These judgments are such that reasonable people may differ and settling differences is what politics is about. This judgment will be discussed further below in the context of uncertainty.

The following are threats to internal validity: history, selection, maturation, instrumentation, testing, statistical regression and experimental mortality (Campbell and Stanley, 1963). Before and after project comparisons are cheap but cannot rule out various explanations for changes in impact other than the presence of the project. An analysis that runs in terms of a comparison between with and without is also needed. Various quasi-experimental designs can control for various threats to internal validity and constitute an academic specialization in itself. These include time series analysis, dynamic control groups, regression models, and pre-project and post-project control groups.

All quasi-experimental designs require specification of variables. This specification can be relatively straightforward in an agricultural irrigation project where we know the factors that must be isolated (controlled) to separate the water input from other inputs affecting crop yield. But these production functions are less well understood for many projects such as education and health. The relationship between inputs and some final impact on human wellbeing may be so poorly understood that there is no choice but to estimate intermediate outputs and value them even when their more ultimate consequence is not well known. For example, education analysts might be forced to try to value the worth of a one point increase in a standard test score because the production function relating test scores to success in a graduate program and ultimately in skill levels and income of graduates are not understood.

Ex-post estimates of the production function are useful for two purposes. One is for accountability to determine if a project accomplished its objectives. It is also useful as input into evaluating a proposed project. A measure of the relationship between an established project and some change in output or impact is still only one point on a production function needed to evaluate a new proposal. To establish optimal size of a new project, at least several points would be necessary to indicate increasing and decreasing returns. In practice, analysts seldom have a production function and must make guesses on scalar questions.

External validity refers to whether the proposed project is enough like previous ones that the ex-post results can be applied. Often the input-output relationship was established in a pilot or experimental setting. When it is scaled up to field conditions, the results may differ. Project designers are always expressing their individuality and creativity, and often add features that they believe may enhance the project, but may actually reduce the results obtained previously.

3. OPPORTUNITY COST ADJUSTMENTS

In many cases, available prices do not reflect opportunity costs and must be adjusted. The adjusted prices are referred to as *shadow prices* and arise in the context of taxation and tariffs, monopoly, subsidies, foreign exchange, and labor policies. The use of shadow prices is one of the differences between an "economic analysis" appropriate for public decision making and a "financial analysis" appropriate for a private firm.

3.1. Taxation and Tariffs

The first-best tax is a lump sum that does not drive a wedge between demand and supply prices. But such a tax is not feasible and adjustments must be made. The opportunity cost of an input or output depends on whether the project affects the total amount supplied or displaces a previously available unit. If the project output adds marginally to total supply of a consumer good, the appropriate shadow price includes the tax because it is what consumers are willing to pay. But, if the project output displaces a previously produced unit, the appropriate accounting value is the marginal cost of production net of taxes.

This reasoning, however, is disputed by Little and Mirlees (1974). They regard indirect taxes as correcting income distribution. They state, "taxation and subsidization of consumer purchases is a useful and socially desirable weapon of policy. Project planners and economic advisors have no general warrant to attempt to nullify the effects of that tax system (p. 224)." For example, government may wish to reduce consumption of some market goods and increase investment or production of government goods. This tax may be a market corrective, and not a mistake. Treatment of taxes is not simply a technical issue where the analyst can assume that the government has made a temporary mistake and really wants to disregard its policies elsewhere.

If inputs are purchased for the project, the value is net of tax if the input constitutes added production. The opportunity cost is the marginal cost of production. But, if the input is in fixed supply and diverted from other uses, the value is what others would have paid for it including taxes. The question again arises whether the government was just raising revenue or was it trying to

restrict private sector use of the input to obtain some policy objective. Knowledge of supply and demand elasticities is necessary to calculate whether the inputs and outputs represent additions or diversions. In practice, this knowledge is costly to acquire and many agencies ignore taxes altogether.

A tariff is just another form of tax. For goods traded internationally, the shadow price is the border price (f.o.b.) for exports and c.i.f. for imports (Little and Mirlees, Ch. 12).[2] This shadow price reflects opportunity cost of production when imports are substitutes. Again, the intent of government policy must be considered. If project output would lower prices to domestic producers outside of the project who are favored intentionally (perhaps infant industry protection), then the government does not want new output even if it were cheaper at the moment.

3.2.   Monopoly and Economies of Scale

A monopoly price above a competitive equilibrium price is a kind of private tax and the same principles apply as noted above. Where marginal costs of production are relevant, an additional problem arises with decreasing cost industries. Such firms cannot price at marginal cost because total costs could not be covered except by taxes. Consider the evaluation of a public waterway project that replaces traffic on a private railroad. The price charged by the railroad necessarily reflects average cost. The cost saving by serving the diverted traffic with the project instead of the existing railroad is the marginal, not the average cost. (This situation raises a distributive issue in that the remaining users will now face a higher average price since total cost is divided by fewer users.) If there is a projected shift in demand for transport services and the railroad is at capacity, then the added project output is valued at the old price. The expected demand curve is the key information needed.

3.3.   Foreign Exchange

The values of project exports and cost of project imports are influenced by foreign exchange rate policy. In theory, the market for currency should adjust in value so that the value of exports tends to balance imports in the long run. In practice, this equilibrium may not occur. The United States has perennial trade deficits even with floating market-determined exchange rates. Many poor countries aggravate their foreign trade deficits with exchange controls, quotas and tariffs over the objection of the International Monetary Fund. If a country accepts its current income distribution and wants to maximize the value of consumption, the imports used in the project that might have gone to private consumers should be valued at what the good would sell for in the domestic market in domestic currency, although it is higher than expected without controls. Shadow pricing assumes that the government wanted to achieve the

results it would have obtained with a general devaluation. But this raises the question of why government used quotas and fixed rates in the first place. The issue again is whether the government's policy is a mistake or a corrective. Governments want to do more than reduce trade deficits; they want to shape the allocation of access to available foreign currency. Dasgupta and Pearce (1972) state, "The shadow price of foreign exchange thus depends on how increments of foreign exchange will be divided among alternative uses, not on the wishful thinking of the project analyst who perceived (or misperceives) the irrationality of the overall policy framework in which he operates (Dasgupta and Pearce, 1972)."

Foreign exchange policy and monetary policy interact. A country might try to reduce imports by devaluation or quotas, which will raise import prices in the domestic currency. The government may accommodate consumers who want to maintain their consumption with an increase in the money supply. The resulting inflation will not reduce imports as expected. Foreign exchange policy can be defeated by monetary policy. CBA is not independent of macro policy, even if World Trade Organization rules and treaties tend to only look at exchange rates and tariffs.

3.4.   Unemployment, Labor and Wages

Disequilibrium in labor markets is another context in which analysts compute shadow prices to replace nominal prices paid to the unemployed. Involuntary unemployment means that the opportunity cost of labor used on a government project is less than the going wage. If labor produced nothing without the project, then there is nothing lost if put to work on the project. The shadow price is zero. In perfect markets, a decline in aggregate demand would result in lower marginal product of labor and thus lower wages. But in practice, wages are often sticky and the unemployed do not or can not offer to work for less than the prevailing wage. Stiglitz (1987a) argues that information costs prevent employers from hiring labor even when its marginal value product exceeds the wage. Lowering wages may decrease the average quality of labor and thus is not in the employer's best interest.

There is both academic and political debate over the existence and implication of involuntary unemployment. Some prefer to maintain the pressure for deflation rather than increase public spending and projects. Employers who experience sustained demand for their product even in general recessions would prefer lower wages and do not want the government to provide an alternative source of employment. Some object to government spending and borrowing in principle and do not want more projects during recession. Some believe that there is no such thing as involuntary unemployment and if labor will not work at lower wages, then it values leisure highly and deserves what it gets. If

governments want to trade off full use of available labor to obtain other objectives such as reducing redistributive inflation, increasing corporate profits, or increasing its foreign competitiveness, it does not want to have labor priced below nominal wages. Even if a shadow price were used, the implied increase in public projects and aggregate demand could be offset by reducing the money supply (Pearce, 1981, p. 109).

In practice, many U.S. agencies use a zero opportunity cost if a project uses unemployed resources. But some analysts make adjustments for the value of home production and part-time work. The chance of a project to draw from the unemployed pool is a function of the required and available skills and the rate of unemployment (Haveman and Krutilla, 1968). The issue is most contentious in developing countries. Lewis (1972) argued that it would be possible to move agricultural labor to public projects with no reduction in agricultural output. However, Stiglitz (1987b) recommends a price above the agricultural wage for urban projects because urban labor is often already excessive from mistakes in migration to non-existent jobs. This rural to urban migration may cause congestion and volatile political protests. More public projects could exacerbate the problem. These raise fundamental policy issues.

3.5.   Intersectoral Interdependence

The utilization of labor caused by the project is more than that involved in construction and maintenance. A project has indirect effects on suppliers of inputs and the processors of outputs. Each of these in turn has suppliers. An input-output model can trace these linked activities stimulated by the project. The model contains a series of linked data of who buys from whom to produce another unit of output. A row in an input-output table shows purchases per unit of gross output. For every dollar of sales from a given industry it buys inputs from other firms and households. These are the first round or direct requirements and they can use unemployed resources. Inputs are needed in turn to produce these direct requirements and so on. These subsequent rounds are called indirect requirements. The sum of direct and indirect requirements is called an interdependency coefficient and is displayed in the Leontief Inverse matrix. Income multipliers are needed for project evaluation. This so called Type 1 multiplier indicates the direct and indirect income changes emanating from a dollar of increased sales of each producing sector. This income is then spent and the induced production of consumer goods further expands the economy. A Type II income multiplier is a ratio combining the direct, indirect, and induced coefficients. This ratio is a large multiple of the original project output.

The input-output tables are a snapshot of a moment in time and do not necessarily indicate what would happen with expanded output in one sector. If

economists could easily predict supply responses to new demand, much of the economic development problem would have been solved long ago. To allow for frictions and bottlenecks, most analysts only use the Type I multiplier. They do not assume that the economy is perfectly coordinated (Haveman and Krutilla, 1968). Once we have the income multiplier, it must be converted to an employment multiplier. Then the chance of the new activity using unemployed resources nationally and in a region where the project is located must be estimated.

Part of the answer to the question of frictions and bottlenecks lies in the incentives for the potential direct and indirect inputs to be produced. An agency may brag upon all of the activity its project has made possible. But, this activity must be rewarded or it will not happen. For example, a flood prevention dam allows expansion of farming and processing. But these cannot happen without complementary investments in such infrastructures as roads, education, and health. From the perspective of other public agencies, it is their projects that made possible the farming and the returns to the dam. The multiplier can be calculated from any point and not every agency can count the same effects. There is no technical answer to the question of allocating credit for employment multipliers. Just as there is no marginal product for complements, credit for the contributions of complementary investments is necessarily partly political. Evaluation, planning and coordination are interdependent. Economic development is more complicated than just ranking individually considered projects and assembling them into a budget.

3.6.   Regional unemployment

There may be sectors and regions of substantial unemployment even if the total economy is nearly fully employed. The shadow pricing of resources in these regions is problematic. Some argue that it is only a question of time when the resources will move on their own, thus obviating the need to bring jobs to them. Or, a retraining program may be preferred to finding projects that can use the available skills. The culture of nations and regions differ in their attachment to place. Governments that accommodate to place preferences will want to shadow price the unemployed labor. Others will not. Immobility of people may be related to immobility of some of their goods. If a region depopulates, there may be large losses in the market value of their houses and the output of community infrastructure.

3.7.   Summary

All are agreed that decisions should be made in terms of opportunity costs, but whose? The price of labor on public projects cannot be separated from the

objectives of macro policy, which are a matter of political conflict. A government that intends to put downward pressure on wages may purposely create unemployment (or take advantage of it when it occurs). The last thing they want is for the project agencies to have larger budgets as a result of higher net returns when wages are computed at some shadow price instead of the higher market price.

The distributive issues can't be settled in a separate transaction. In a second-best world, the government is likely to want to pay the nominal wage. For example, in the U.S., the Davis-Bacon Act requires it. If projects are built using shadow prices that would not otherwise be built, they are the occasion for some taxpayers to make transfers to project labor.

4.   PRICING BENEFITS AND COSTS

Analysts are called on to supply prices when project inputs and outputs are not marketed. This problem requires inferences from indirect evidence of willingness to pay that is the essence of the economist's technical expertise.

4.1.   Analogous Good Method

One of simplest methods is to reason from the price of an analogous good to that of the non-marketed project good. This approach returns to the first topic above, namely to establish that the goods are perceived as comparable.

A common application of the analogous good method is to projects that increase human longevity. People are in effect purchasing longevity when they buy products that reduce risk or trade reduced income for increased longevity. The inferred value can then be used to value projects that increase longevity. The difference in income in jobs of different risk is a measure of the willingness to pay to reduce risk. The method depends on the analyst's ability to control for other factors affecting income differences (Freeman, 1979).

4.2.   Intermediate Good Method

Even though the output of a public project is not sold, the output may be an input into the production of a good that is sold. The method requires the estimate of an enterprise budget with and without the project. It requires data on the production function including all inputs and their prices, and the price of the final non-project product. Project gross benefit equals the change in net income. It is what the user of the project output could afford to pay for an input into the user's activity. A flood control project is a common example. Flood control is a high exclusion cost good and cannot be marketed directly because of free riders. But flood control changes net income from farming in the flood plain. The method is also widely used to evaluate education and

training programs. The benefit is the change in net income of the participants with and without the program (Hardin and E. Borus, 1971),

A human capital conception is commonly used in valuing life in the context of safety projects and regulation. Life can be viewed as input into income production. Mortality and morbidity reduce lifetime earnings. The health project output may not be marketed, but it is input into labor that is marketed. A number of political issues arise. Should consumption or total income be the correct measure (Jones-Lee, 1976)? Should the measure be based on objective historical data or individual subjective estimates of participants (Akehurst and Culyer, 1974)? Should all impacts be monetized (Viscusi, 1996)?

To use the opportunity cost of lifetime earnings as a measure of project output value is to make a political choice of property rights and income distribution (Kelman, 1981). The practice includes a decision to put the potentially harmed person in the position of a buyer of safety rather than a seller entitled to be free of harm. This approach raises the first-best vs. second-best question. If income distribution had all been settled or one could obtain the desired distribution outside of projects and regulations via costless transfers, then the human capital approach would be unambiguously Pareto-better. The same point can be made with respect to environmental products.

The value of the intermediate project good can be derived from linear programming which estimates a shadow price for any input whose use is constrained bellow the optimum level. For example, if farmers are constrained in fertilizer use because of concern for ground water quality, a linear programming model will estimate the reduction in income (value lost because of an inferior input substitution). The income lost is the opportunity cost of regulation to be compared to its benefits.

4.3.   Cost Saving Method

A current expenditure is evidence of willingness to pay. If the project can reduce expenditures that would otherwise be made, this is a benefit. If the project is a perfect substitute for a former expenditure, the saving is a measure of gross benefit. A common application is in transportation projects (Harrison and Quarmby, 1974). A project may substitute for an existing mode of transport and save fuel, time, wages of drivers, repair to equipment, and capital costs of goods in transit. Another application involves damages avoided and reparative expenditures. If there is effective demand for the repair, then its avoidance is a benefit. Medical care is an example. The prevention of an injury for which medical care would be required is a benefit.

Differences in human perception create the need for political resolution. For example, the cost saving method commonly used in transportation project evaluation requires a choice between the analyst's perception of time saved and

that of the actual users of the transportation. The same problem arises in the context of exposure to hazardous events. Sugden and Williams (1978, p. 179) give the label of "merit goods" to products that people would want if they understood their best interests. Whether this approach represents desirable caring or paternalism requires political judgment.

4.4.   Access Cost-Quantity Demanded Method

Another opportunity to infer the value of project output is when consumers face different market-valued costs of access to a non-marketed good. Travel costs to gain access to recreational goods is an example and the method is often referred to as the travel cost method although other access costs are possible (Smith, 1971) (Bockstael, 1995). The method works best when applied to a specific site that is the main purpose of a trip. The travel and time costs are estimated for different distance zones and the rates of attendance from each zone gathered. It is then assumed that if a person with no travel cost were asked to pay an entrance fee equivalent to the costs faced by others, their quantity demanded would be similar. It must be assumed that people respond to the total cost regardless of its composition. Statistical models can be developed to control for income and other socio-economic variables thought to affect demand (Gum and Martin, 1975).

The conversion of hours spent to dollars is a problem if there is disequilibrium in labor markets. The wage rate may not indicate the marginal value of time or money for people with all or nothing fixed hours (Bockstael, Strand, and Hanemann, 1987).

4.5.   Rent and Hedonic Price Models

In the above discussion of access cost-quantity demanded method (travel cost), value is inferred from different behavioral responses to variations in the cost of a market valued good controlling access to a publicly provided good. It is also possible to compare the prices paid for different goods with differential access to a publicly provided good. For example to benefit from an environmental amenity, one must pay more than for an otherwise similar good without the amenity provided by public investment or regulation. A house overlooking a public park will have a higher price than one of similar size some distance away. This economic rent (return above opportunity cost) is direct evidence of willingness to pay. For a producer's good, it is the capitalized present value of the flow of expected future income from the better situated land.

Valid measures require that the market be competitive and buyers mobile. This requirement will insure that no user surpluses will exist when a new equilibrium is reached after implementation of the project. Then any increase in utility or factor returns gets translated into observable differences in rent. For

example, if air quality were improved in a slum area and no new buyers were available to bid up the prices, then utility perhaps increased but it is not observed in market prices.

Regional boosters often claim large benefits from a project that attracts new migrants and business. They cite increases in land values around the site but ignore changes in rents off site. Only the *net locational advantage* of a site contributes to national income. For example a business may move and thereby decrease rents in its old location and increase rents in the project area. It is only the change in the business net income with and without the land enhancing project that is a benefit, and not the before and after rents at the project site. Freeman concludes "in general, property value changes can be interpreted as benefits only when there is some mechanism to assure that there are no economic surpluses accruing to households, and when there are no changes in wages or other factor prices (Freeman, 1979, p. 151)." Also see (Hoehn, Berger, and Blomquist, 1987).

In many cases, land rent is embedded in the price of a multi-attribute product. These other attributes can be controlled in a *hedonic equation* and calculation of the implicit price of the project produced attribute. For example, to estimate the value of a project that enhanced air quality the value of housing (X) is related to air quality (Q) and a number of other variables such as number of rooms (C_k) and lot size (C_j). The implicit price function is:

$$Px_i = f(C_{ik}, C_{ij}, Q_i)$$

The implicit price of the project enhanced characteristic can be found by differentiating the implicit price function with respect to that characteristic. This differentiation gives the increase in expenditure of (X) that is required to obtain a house with one more unit of air quality *ceteris paribus.* See (Harrison and Rubinfeld, 1978) and (Freeman, 1993, Ch. 11).

4.6.   Contingent Valuation

Instead of inferring values by past choices, people can be asked directly of their willingness to pay. This method is the only alternative for new products. People can be asked their maximum willingness to pay for various quantities of a good (Cummings, Brookshire, and Schulze, 1986) (Bateman and Willis, 1999). The sample data can be projected over a relevant population of users of the project and a demand curve estimated. This approach is often called the contingent valuation method because the values obtained are contingent on the interpretation of hypothetical markets and products. The responses are known to be affected by choice of the survey format, reminders of alternatives, starting place bidding values, product description and familiarity.

As these individually constructed studies accumulate and are put before decision makers, there is a question of additivity and whether the consumers budget constraint is fully operative (Brown and Shogren, 1998, p. 12). In theory, consumers always have in mind all the ways they might spend their income. But in actuality consumers are affected by the sequence in which items are called to attention. In grocery stores, sales of an item can be affected by placement in the store. It might be referred to as the "end of aisle phenomenon." This phenomenon is no less true in politics.

The use of contingent valuation (bidding games) requires the resolution of political questions.[3] The process can't begin unless it is decided whether the question is willingness to pay or sell. As noted in another context above, this is a basic property rights question that is antecedent to market exchange or any simulation thereof. The framing of the questions, the anchor point, and the degree to which opportunity cost tradeoff is made explicit all are known to affect the resulting prices. Legislators seek membership on the rules committee because control of the agenda affects outcomes. Economists at least since Kenneth Arrow (1963) also understand that grouping of issues (order of vote) affects the formulation of winning coalitions. Yet, this is ignored when analysts go off by themselves and make an independent contingent valuation study without political input.

The parallels between surveys (contingent valuation) and a politician sampling and acting upon constituents' preferences is striking. Both processes are subject to the same issues of sampling, framing, and aggregation. Both involve issues of property rights. Some economists are willing to rewrite the constitution and promise to independently measure the revealed "true values" of sovereign voters. Others say, "there may be no single 'true' behavior if preferences vary across time and between choice-making circumstances (Shabman and Stehenson, 1996, p. 441). The question of true value is not simply a problem of principal-agent. Rather, the issue is the aggregation of the preferences of multiple principals and the dynamic learning environment of the principals whose preferences are evolving. After reviewing the effect of alternative framings of willingness to pay questions, McFadden (1994, p. 706) concluded that "The experiments display patterns that are more easily explained by 'constructed' preferences rather than by rational individualistic stationary preferences." An extended argument is not possible here, but there is nothing inherently superior about market prices or prices inferred from indirect evidence of willingness to pay or surveys vs. administrative prices, i.e., reservation prices set on publicly owned resources or bid prices to acquire resources for public use (Schmid, 1989) (Vatn and Bromley, 1994). If it is legitimate for the legislature to change property rights and generate alternative prices in the market, it is legitimate to choose those prices directly under one constitution or set of political rules (or surveys) or another.

4.7.   **Existence and Option Benefits**

The methods described above are applied to the active use of goods. People also value the passive use of goods. They derive utility from just knowing that something exists. For example, people may derive satisfaction from knowing that giant redwoods exist even if they never plan to see them. This is referred to as *existence value* (Krutilla, 1967).

People may value an option to utilize a good. Just as investors buy a stock option that they may or not exercise, people value a project or regulation that preserves an option to enjoy a product such as the environment, a hospital, or a transportation alternative. Further, it may be desirable to delay taking an action that would be costly to reverse. The term *quasi-option value* can be used to describe differences between the expected value of delaying an irreversible decision and the value of adopting it immediately (Arrow and Fisher, 1974). Purchase of this option allows for learning and changing preferences. Contingent valuation is typically used to measure existence and option benefits (Vining and Wimer, 1998).

5.   **NON-MARGINAL PROJECTS**

Some projects and regulation are large enough to cause a change in prices of the output and inputs. A non-marginal project may depress prices such that it has no net benefits. Yet, consumers are better off. Many applied economists measure this welfare change with an estimate of the *compensating variation (CV)* defined as the maximum willingness of an individual to pay that would keep them on the same utility level as <u>before</u> the project. Alternatively, the welfare change could be measured with the *equivalent variation (EV),* defined as the willingness to accept an amount that would maintain the same utility level as obtained <u>after</u> the project. It is the income necessary to forgo the benefits of the price decrease. These measures need not be the same (Hanemann, 1991). They differ because of income effects, substitution effects and loss aversion, CV is limited by a person's income while EV is not. Mishan (1976) advocated that a project must have net benefits using both measures.

EV and CV require holding utility constant and are difficult to measure empirically. Consumer surplus (CS) approximates EV and CV and is more tractable, but has its own conceptual problems. Consumer surplus is the amount that a consumer would pay over actual payment. It is represented by the area under the demand curve and above the price (the total area under the demand curve for a non-marketed good, which is then compared to cost of the project). Empirical measurement is complicated to account for changes in the prices of multiple other goods, path dependence as a function of the alternative sequences of change in the various prices of substitutes and complements,

adjustments in the quantity demanded if maximum willingness to pay is extracted for intra-marginal units (compensated demand curves), income effects, and multiple period analysis to account for learning, and consumer adjustments. Slesnick argues that these and other considerations are theoretically, conceptually tractable for the individual consumer, but there remain severe data problems in practice. Use of household data requires a "large number of observations on household demand that include information on the prices paid for the goods … further application of these methods will require the parallel development of alternative data sources (Slesnick, 1998, p. 2124)."

The use of consumer surplus for project justification is equivalent to a firm being a perfectly discriminating monopolist. The amount of consumer surplus that such a firm can extract depends on whether other firms are also trying to do it (Hoehn and Randall, 1989). Not all firms, whether private or public, can simultaneously extract the consumer surplus that is estimated for each one acting alone assuming no one else is trying. As Samuelson (1963, p. 197) once noted, some otherwise bankrupt firms would have survived with price differentiation.

Even if the conceptual and empirical problems of measuring CV, EV or CS for an individual are solved, it is quite a different matter to aggregate them into a measure of social welfare. This question is addressed in the social welfare section below.

6.   VALUATION OVER TIME

Projects have different cash flows over time, differ in size and length of life. To compare projects there must be a common denominator. It is the capital market that allows any particular cash flow (with negative and positive values) to be converted to reference standard. One can borrow to bring consumption forward or lend to delay it to the future. A short project can be made comparable to a long one by reinvesting the proceeds. The critical element to this process is the rate of time preference.

A unit of value can be consumed now, or invested and consumed later. If it is consumed now, it may be inferred that the utility now is greater than the utility later even if the number of units is greater. There is always a reference rate of transformation of present to future values. It may be the rate of interest on a savings account or bond. If the future value is rejected, it implies that the rate of growth in value produces less utility than present consumption. The project competes with consumption and other investments. The project's ability to transform today's income into tomorrow's is compared to the utility of consumption and alternative investments. So if a person has a time preference of 10%, that person will prefer consumption or alternative investments unless the project can transform today's income (project cost) into future consumption at

a rate greater than 10%. Tomorrow's cash flow is discounted by the rate of time preference. The rate or speed at which the future consumption is discounted backward can be expressed as $(1/1+r)^t$ where r is the rate of discount and the exponent t is the year in which income is received. If the discounted value of the project is less than one, it is rejected. Alternatively, one can think in terms of terminal values of the longest available investment. The project (or alternative investments) is transforming or compounding present values into future values. Unless the project can do a better job of compounding, it is rejected. Compounding or discounting at the same rate of time preference is a symmetrical process. An amount compounded forward to a terminal value, and then discounted at the same rate produces the same value starting place. (More on investment criteria below.)

6.1.   Differences Among Individuals

A person's time preference varies with their present and expected wealth and its certainty, age, impatience, and concern for future generations. Generally we expect poor people to have a higher time preference than the rich. Still, if all had access to the same capital markets and could borrow or lend at the same rate, they would have the same *marginal* rate of time preference. A person with the higher rate will be a borrower and the person with the lower rate will be a lender until their rates converge in general equilibrium. In practice, capital markets are not perfect. People cannot lend and borrow at the same rate because of risk and transaction costs. Lenders may ration credit. Different amounts of savings earn different rates of interest. Individuals face different tax rates that are not proportional to benefits. Differences in marginal rates of time preference can persist and government will have to choose between rates preferred by different people.

6.2.   Cost of Capital Approach

Differences in time preference among individuals would not matter if government made decisions according to its cost of capital. As long as the government project has returns greater than the cost of its capital, it can make people with different time preferences better off. However, they will differ as to the means of financing (Sugden and Williams, 1978). Assuming a project is to be undertaken, persons with a marginal time preference lower than the government's cost will prefer to be taxed since they do not have as good an opportunity as the government. Persons with a higher marginal rate will prefer government borrowing since they have better things to do with their own money. Further, individuals with different rates will differ over project ranking because they would want to use different rates for reinvestment compounding to compare projects of different lengths.

Nevertheless, if the cost of capital is what policy makers want, what is it? Some suggest it is the after-tax rate of return on long-term government bonds (Lesser and Zerbe, 1998, p. 256). The length of term is usually left vague since projects are of differing lengths. Is this to be the nominal rate at the time a project is evaluated or the expected rate? If it is the expected rate, whose expectations count? In principle, if the benefits are in nominal terms then the discount terms should also be nominal. In practice, it is sometimes difficult to tell whether the projected future prices an agency uses are in real or nominal terms. There is a huge literature of differing estimates of the opportunity cost of capital. Lesser and Zerbe examined the rate of return on government bonds since World War II, on low-risk railroad bonds, and on commercial paper from about 1887 and conclude that 3% was the cost of capital in inflation adjusted dollars. (Lesser and Zerbe, 1998, p. 262).

In practice, the U.S. water resources agencies use the nominal rate on 15-year government bonds in the year previous to a project's evaluation. Changes in this rate are slowed by limiting change in any one year to 0.25%. Other agencies use different rates or none at all.

Is the discount rate for public projects a matter of data to be observed, or a public choice to be decided and given to the analyst? With a perfect capital market, everyone would have the same time preference at the margin. People with initially different time preferences would borrow and lend, and market rates would adjust until all players are in equilibrium. But in disequilibrium, people have different opportunities and differ over the desirability of financing public projects by borrowing or taxation as noted above. Some further political resolution of conflicting interests is necessary even if the distribution of factor ownership (wealth) were acceptable and there were many borrowers and lenders (no market power). The literature, of which Sugden and Williams (1978) and DeAlessi (1969) are representative, seems to have shifted to viewing the choice of discount rate as a political decision rather than a datum to be discovered. Pearce and Nash (1981, p. 164) observe that "no single school of thought on discount rates commands consensus among economists … the issue is one of choosing a discount rate in a second-best world, so that behaving as if first-best conditions prevailed … does not seem relevant."

6.3.   Inter-generational Discounting

Inter-generational trade-offs in use of the environment are especially marked by differences in preferences. The property rights issue is not just differences in time preferences among members of the present generation, but which generation owns the environment. Choosing a low discount rate does not clearly favor future generations because of its opposing effect on intensity of resource use and the scale of total development (Pearce, Markandya,

and Barbier, 1989) (Daly, 1991). If future generations are to be given rights, it is best achieved by some sustainable development constraint rather than by discount rate policy. This argument is the rights equivalent of making future generations joint owners and then government as its agent (trustee) deciding not to sell to the present generation of users. None of this political question is instructed by technical measures of any existing discount rates. Robert Solo suggests that "Maybe the idea of a unitary decisionmaker—like an optimizing individual or a wise and impartial advisor—is not very helpful when it comes to the choice of policies that will have distant-future effects about which one can now know hardly anything. Serious policy choice may then be a different animal, quite unlike individual saving and investment decisions (Solo, 1999)." Solo uses the term "responsibility" to describe the choice context. This issue raises questions of deontological rights rather than willingness to pay (Adler and Posner, 1999). It is about ideology and working out the meaning of doing the right thing.

7.   INVESTMENT CRITERIA

There are several investment criteria or tests for project acceptability and ranking. They differ in their implicit assumptions on reinvestment of net cash flow and the manner in which the different scales of projects is accounted for. When these dimensions are standardized, all criteria produce the same results.

7.1.   Net Present Value

NPV is the summed discounted value of the cash flow produced by the project. $\text{NPV} = \sum B_t / (1+r)^t$ where B equals the net cash flow in each year t. It could be zero, negative or positive. The rule is to invest in any project with a positive NPV. It would be efficient to expand the budget so that all such projects could be undertaken. If for some reason there is capital rationing, it will be necessary to rank projects. This ranking should not be done by going down a list of projects and stopping when the budget is exhausted. Since NPV is an absolute amount this would give an advantage to large projects. For ranking, some sort of rate is needed to relate return to a unit of capital, or the NPV must be maximized for some set of projects that exhaust the budget.

The opposite but symmetrical concept to NPV is net terminal value (NTV). It uses compounding instead of discounting. $\text{NTV} = \sum B_t (1+r)^t$. A project with a positive NTV will also have a positive NPV.

7.2.   Internal Rate of Return

The IRR is defined as the rate of discount that reduces the cash flow of a project to zero net present value. It is the discount rate r for which the sum

of $B_t(1+r)^t = \text{zero}$. It is an average rate of return. A project that produces a given IRR is equivalent to the time flow of the given initial investment compounded forward for the given number of years at an interest rate equal to the IRR. The choice criterion is to invest in any project in which IRR is greater than the opportunity cost of capital. The World Bank uses IRR to qualify projects for loans (Gittinger, 1982, p. 331).

7.3.   Benefit-Cost Ratio

A rate of return can also be constructed out of present values. A common measure is to compute the ratio between the present value of benefits and the present value of costs. Any ratio involves a choice of denominator, usually what is considered to be the most limiting source of funds. This choice is a matter of political judgment. Capital (K) is defined as negative cash flow in any year. Operating cost that can be covered by receipts is not limiting and can be netted from the numerator. Still, Eckstein (1961, pp. 63-64) observed that U.S. agencies receive an annual appropriation that must cover both capital and operating outlays. Nevertheless, the clients of these agencies may not want consider operating costs as limiting. A more serious problem is that if operating costs are not considered limiting it will encourage agencies to classify as many costs as possible as operating when they design the project (Kuhn, 1962, p. 174). There is also a question of the source of capital in federal systems where part of the cost is from a local government and part national. A ratio could be constructed to reflect a political judgment on whose capital is the most limiting. There can also be an issue of whether to count as limiting private investments that complement the public investment.

7.4.   Criteria Comparisons

Use of any investment criterion makes an assumption about reinvestment of net cash flow. Reinvestment is necessary to compare projects of unequal life. A net present value calculation carries forward net cash flow indefinitely at the rate of discount chosen to calculate NPV. The internal rate of return carries forward net cash flow at the internal rate of return implicit in the project. These rates are unlikely to be the same and so the two criteria can produce different rankings though they will both identify the same set of projects with some net return. This inconsistency can be eliminated by the *terminal value method* that makes explicit the rate of reinvestment (Schmid, 1989, p. 205). If there is not another project that earns as much as a short lived project, then the latter's IRR is not relevant. Likewise, if there is another project that earns more than the rate used to discount cash flow, that rate of discount is not relevant.

If the cash flows are explicitly compounded at the politically chosen opportunity cost, then all criteria will produce the same ranking. Any political judg-

ment on the opportunity cost of capital reflecting disequilibrium can be incorporated. For example, the benefits received by those with better opportunities than the government will want to receive the net cash flow and make their own reinvestments while those with poorer opportunities will want the government to reinvest if it has better alternatives. Different opportunity costs for portions of capital could be implemented if politically desired. This standardized and reconstituted cash flow reflecting reinvestment can produce a Normalized Terminal Value Ratio with whatever definition of capital is desired. If preferred, a normalized IRR could also be computed. Once the different assumptions implicit in each criterion have been superceded by transformation (actual cash flow over a common period) all criteria will produce the same ranking (Mishan, 1976, Chs. 37 and 38) and (Robison and Barry, 1996, Chs. 4 and 6).

The ranking issue may be of little importance in practice. Ranking is not used for decision making in the U.S. or by the World Bank. No politically favored project is ever forgone just because a higher return one is on some list. No agency publishes a ranked list of projects. Neither does the U.S. executive office publish a list of ranked projects across agencies. Since all criteria described above identify the same list of projects with some net returns, the ranking issue is moot. This equivalence would not be the case if the rate of return were used to size projects. If different sizes of projects were explicitly considered for a site or purpose, then their ranking would be affected by different criteria. In practice, project designers often have some rule of thumb for what a project should look like. Where project size and capital intensity are explicitly examined, there is a tendency to extend scale so that the net benefit of each project is maximized where marginal cost equals marginal benefit. This approach is incorrect with capital rationing. The scale should be extended until the ratio of marginal cost to marginal revenue is the same as that ratio for the marginal project (which will be earning more than the opportunity cost of capital with capital rationing).

8.   UNCERTAINTY

Project outcomes are often uncertain. History is full of examples of project failures and cost overruns (Pohl and Mihaljek, 1992). Future prices of outputs and operating inputs are problematic. There are some projects in which it is possible to attach mathematical probabilities to outcomes with some confidence. In other cases, expectations can only be qualitative.

8.1.   Expected Value and Expected Utility

The different possible environmental conditions that affect outcomes are referred to as states of nature. These might be physical conditions such as the weather or other things affecting the production function or prices. Take

the case of weather that would affect the outcome of an agricultural project. Fifty years of rainfall records would allow analysts to state the probability of a range of rainfall amount. There is an unavoidable element of subjectivity even when some data are available. One hundred years would be better. There is uncertainty about the probability estimate. The range of possible outcomes and their probabilities can be described in terms of expected value (EV) which is the weighted average of the outcomes. $EV = \sum_i p_i x_i$ where x_i is the value of the ith possible outcome and p_i is the probability of its occurrence. The expected value is an appropriate description of cash flow if the decision-maker only cares about the mean and not its variance. Variance is equal to $\sum_i p_i (x_i - EV)^2$. However, the project with the highest expected value may not be preferred to another if its variance is high. For a detailed exposition of risk preference measures see (Zerbe and Dively, 1994, Chs. 15 and 16).

A person who is prepared to act upon expected values regardless of variance is said to be *risk neutral*. The risks must be independent, and future states of nature should not have wealth effects or threaten survival. The risk neutral are in a position to wait as needed for the average result to occur. But if there are costs to risk bearing such as going bankrupt before the results average out, such a person *is risk averse* and is going to trade off mean and variance.

The tradeoff of mean and variance can be made formal with the concept of *expected utility* (EU). This tradeoff is implicit in a utility function relating income to utility and would incorporate any diminishing marginal utility of income. Such a utility function might be revealed with a bidding game asking people to choose between different payoffs with different probabilities. It is necessary to assume that utility is proportional to probability. While this is not absurd, there is no compelling reason to think it is true for most people for most kinds of projects.

Can the government be regarded as a risk-neutral investor applying the expected value criterion? The government can in principle spread the risk across many people, thus making each person's risk small (Arrow and Lind, 1970). But in practice there is no institution for adjusting taxes to variation in the states of nature and to distribute benefits as dispersed national dividends in proportion to taxes. Groups lobby for projects because of unique and large expected benefits, and if disappointed, they have used up their political capital and may not get another project. Thus they are not risk neutral. It cannot be assumed that people want government to make risky decisions as if variance and beneficiaries did not matter.

People have different preferences for the tradeoff of mean values and their variation. There are few markets for recording these preferences. Each person can't independently adjust a portfolio of public projects to obtain their preferred risk exposure. This difficulty means a political judgment is needed.

Uncertainty is an area where behavioral economics has a lot to offer. Greater risk is tolerated if that risk in voluntary, immediate, known precisely, controllable, and familiar. Much of the decision analytics separates the perception of mean values from perception of their variation, but much empirical evidence exists that the perceptions are inter-related. The frame for viewing attitudes toward uncertain events has a lot to do with what is seen. A political compromise among differing perceptions is needed.

In practice, government employs a variety of directional rules of thumb rather than any formal all-encompassing formula. Some agencies do use expected values, some add a contingency allowance to costs, and some make the case that their estimates were a conservative choice from among alternative estimates. Benefits and costs often turn on population and utilization projections. The methods used to estimate future prices are sufficiently vague that it is difficult to tell what assumptions are being made. The only thing that can be said is that there is some method that the agency and its clients have become comfortable with. No matter how careful and systematic other dimensions of analysis may be, adjustments for uncertainty are sometimes necessarily so loose that they may overwhelm the implications of other judgments. It is common for an analyst to offer a sensitivity analysis. This analysis indicates how project outcomes might differ with various states of nature, but is in itself not a systematic choice tool.

With some kinds of projects there is *fundamental uncertainty* and it is not possible to even name the categories of possible effect let alone their probability. Perhaps the implication of all of this is for management and monitoring of whatever projects get implemented, rather than worrying about which project to build. In an uncertain world, flexibility is a virtue. Designing into the project an ability to adjust as the future unfolds may be more valuable than a formal process for comparing projects of fixed dimensions.

9.   SOCIAL WELFARE

What can cost-benefit analysis say about social welfare? It has become popular among applied economists, particularly in recreational and environmental projects, to sum individual CV's or EV's or consumer surplus as a measure of the social value of a project including those with a price decrease. Most theorists on the other hand have been hesitant to endorse it. The beneficiaries of a project may not be the taxpayer who pays for it. A project may increase the income of the rich and decrease the income of the poor. The problem is that of the first-best assumption of optimal income distribution in the many-consumer economy (Tresch, 1981, p. 198).[4] If costless lump-sum transfers are not available, Boadway and Bruce (1984, p. 271) conclude that, "The use of the unweighted sum of household compensating or equivalent variations as a

necessary and sufficient indicator of potential Pareto improvement is rife with difficulties." The theory has been worked out mostly in the context of taxes but the implications are the same. Tresch (1981, p. 351) says "it may not be very useful to think of the effects of distorting taxes in terms of deadweight loss. Unambiguous notions of efficiency loss involve the use of the expenditure function, which is best suited to one-consumer economies."

A review of the literature led Slesnick to conclude that "it is now widely accepted that consumer surplus should not be used as a welfare measure (1998, p. 2108)." The restrictions necessary to use a summation of individual CV's or EV's include homothetic preferences (all income elasticities of demand equal one) or preference functions must be parallel with respect to the numeraire good to avoid path dependence (Chipman and Moore, 1980). This requirement is contrary to empirical evidence. Another restriction is that all individuals have the same marginal utility of money. Again, this seems contrary to evidence. Assuming a representative agent will not do.

The Pareto Principle is often used to finesse the problem of interpersonal welfare comparisons. A good project has net gains such that the gainers could compensate the losers and still have something left over. (Some are better off without anyone being worse off.) This unanimity requirement seems unreachable in practice. Perhaps then the test could be only that of a potential Pareto improvement. The project would be acceptable if the gainers could compensate the losers even if not actually done.[5] Choice of test would seem to require a major ethical judgement. Sen (1979) argues that the New Welfare Economics is irrelevant. If compensation is not actually provided or could not be provided with lump sum taxes and transfers, it has no implications for public policy.

The point is not that the Pareto Principle or potential Pareto criterion is right or wrong, but rather that *any* criterion is a political value judgment. As Slesnick (1998) puts it, "Any effort to develop an index of group welfare must inevitably make normative judgments in which gains to some are weighed against the losses to others (p. 2137)." The problem with aggregation "is that the expenditure function provides an exact representation of individual preferences for a fixed set of reference prices p. The choice of the prices used to 'cardinalize' preferences needs to be invariant to this choice (p. 2141)." While use of CV or EV "has the appearance of being a positive measure of the change in aggregate welfare, it is no less normative than methods based on explicit social welfare functions. The sum of Hicksian variation depends on the distribution of well-being and the underlying ethical assumptions are often ambiguous (p. 2151)." Any use of the Pareto Principle privileges the *status quo* which requires an ethical judgment. And even what constitutes the *status quo* is a matter of interpretation and judgment.

The applied economist can scarcely do better than conclude as did Little (1957, p. 184) when he said "The best criterion for investment decisions must,

within wide limits, be determined at dynamic and administrative levels—and not at the level of static welfare theory." Which private or public firms are to be allowed to act in terms of being a discriminating monopolist is fundamentally a distributive question. Consumers don't know what they are sovereign over until they know whether they will be subject to price differentiation, and firms producing complements and substitutes are not sure of their prices until they know whether other firms can differentiate. Note that if you approve of the distributive implications you call it price differentiation, but if not, you call it discrimination.

The constitutional and institutional context for CBA is implicit when Ball says, "There are an infinite number of possible Pareto-efficient points for the economy of the neoclassical model. Attainment of any of them depends on the initial distribution of resources between individuals in the economy. For each of these infinite number of Pareto-optimal points, there consequently corresponds a different set of 'efficiency' prices (Ball, 1979, p. 76)."

Benefits and costs are what they are because of the underlying income distribution and the distribution of property rights (Samuels and Schmid, 1997). "Benefit-cost analysis is inseparable from the law in important respects. That is why a benefit-cost analysis does not make the decision itself. Rather, benefit-cost analysis requires the law to make clear the pattern of major rights that frame the analysis, and to determine whose values are to be counted (Lesser and Zerbe, 1998, p. 241)." The contrast of EV and CV is a function of property rights. The starting place for EV is that of a rights holder listening to bids while CV is that of non-owner who is a buyer rather than a seller of an opportunity. Like all property rights this must be legitimated by governmental collective choice. The analyst must be instructed and can not arbitrarily choose the starting place.

Whether one uses the sum of CV's, uses a social welfare function or just sums real income changes, one is making an ethical choice. For example, in a health project, the life of a poor person measured by willingness to pay will be a function of future income preserved. This measure will make saving the life of a rich person more valuable than that of the poor. In practice, EPA uses the same value for all persons to justify a regulation (Adler and Posner, 1999, p. 86). Likewise, in a flood control project, saving a rich person's house is more valuable than that of a poor person. In practice, the Corps of Engineers does just that. They get signals from Congress as what is acceptable. Congress would prefer that the project meet the net benefits test without any explicit weighting or income distribution objective. The fact that the project does not rank high because a large portion of the beneficiaries are poor, is no problem since ranking seldom plays a role anyway. If the political support is there to help the poor in a particular location, the project will be funded over

another with higher total benefits from whatever source. Analysts point out apparent inconsistencies in the amount spent to save a life in various regulatory programs (Tengs and Graham, 1996). Is this unsystematic or does the political decision regard these lives as different? Interpersonal comparisons do get made.

If analysts cannot provide an independent aggregative measure of social welfare, what can they do? Sen is at the frontier of thinking about social choice. He advocates a "capacities" or "functionings" approach to get around all the problems with income measure (Sen, 1985). If income and consumption are treated as an input to such capabilities as the degree of personal liberty, level of justice, life expectancy, health, etc., then we have something closer to a measure of utility. All of the money-metric problems such as reference prices are avoided. Slesnick believes that Sen's approach is theoretically sound, but alas, "It is difficult to imagine how this approach might be implemented empirically to provide a comprehensive welfare measure. Individuals' capabilities are not always the result of revealed preferences so we have little prospect for measuring individuals' valuation of their capabilities" (p. 2148).

Perhaps measurement issues miss the point. The issue is not how to sum up value choices already made (i.e. revealed), but how to inform their creation. Is not the relevant question how to structure public debate and political choice? Shall analysts ask the politicians for weights to be applied to the income of different people?[6] Or shall we ask them to set substantive goals, for example, levels of living or greenhouse gas abatement? For example, Kopp and Portney suggest mock referenda for intergenerational decision making to get around the long term discounting problem (Kopp and Portney, 1999).

Sen's capabilities at the very least tie back to the choice of input and output categories and the taxonomy of products. The choice of product names and features is informed by consideration of what makes it good, i.e., how it serves utility. Marketers of private goods in their advertising try to tie their product to such things as sociability, re-creation, health, and security. When the analyst describes a good for contingent valuation, implements the analogous good method, or constructs a hedonic regression equation with variables to describe a multi-featured marketed good, something like Sen's capabilities and functionings are implicit or explicit.

10.   CONCLUSION

Analysis of each of the steps in cost-benefit analysis identified above reveals conflicts of interest (preferences) which can only be resolved by political (collective) choice of property rights assigning opportunities to the various parties. The rules of CBA function as property rights equivalent to private property rights established by legislative and court decisions for the market economy.

A simple political weighting of the traditionally computed benefits received by different groups does little to clarify and inform the debate necessary to resolve the disputes over the several specific varieties of rights conflicts.

The separation of technical analysis and political choice is not tenable. Theory and experience point to a more interactive, iterative relationship between analysts and politicians. "The role of benefit-cost analysis is not to make decisions, but rather to inform them by providing information relevant for decision makers (Lesser and Zerbe, 1998)." The distribution of rights affects prices and thus any existing set of prices can't guide the choice of rights including those embedded in rules for appraising public spending and regulations. It can't be assumed that political choice has once and for all chosen the distribution of rights and that the only problem is the technical one of implementation of the preferences of rights holders. The process of public investment and regulation is never wholly exchange facilitating (solving market failure) nor wholly grant making, and the distinction needs continuous political input. The analyst need not apologize for asking more questions of the politicians. And the technical input is no less useful for the fact that as new politicians are elected, public investment priorities change. The value of analysis is in clarifying the substance of evolving preferences and in being explicit about whose preferences count. Mishan (1988, p. xiii) argues that the essence of an economic approach is the "basic maxim that individuals' expressed valuations alone are to count." But it is the responsibility of collective action to decide which individual counts when individuals conflict. The question of who counts cannot be escaped whether the analyst uses a financial analysis containing only available and nominal market prices or an economic analysis containing shadow prices and non-market evaluations.

Which politicians do the analysts ask to obtain the necessary property right choices? Wouldn't an independent technical analysis be better than incorporating the policy objectives of a corrupt and unrepresentative government? Each analyst must make his or her own moral judgment. But, however tempting it is to right a perceived wrong, ultimately all must ask "who does the analyst represent and who elected the analyst to this job." Just as there is no market without property rights, there is no public choice without a constitution (civil rights). Alternative constitutional rights give more weight to some interests and less to others. The full implications of Arrow's Impossibility Theorem must now be acknowledged. There is no aggregate public interest which a political system can be designed to reveal with fidelity any more than there is an aggregate consumer interest which an economic system can be designed to reveal. Distribution of opportunities is a matter to be argued, chosen, and worked out, not just something pre-existing to be revealed. The debate can be informed but not instructed without presuming the answers to the distributive questions. Efficiency in public expenditures is not a prior fact out there waiting to be

discovered, but is an artifact to be worked out. CBA is an information input into that political process and can make it as systematic as bounded rationality allows.

Nothing here should be read as a critique of CBA. The point is that CBA and economic efficiency have a constitutional base. CBA or any alternative requires political, ethical judgment and the key issue is whether these are hidden as technical issues or are explicitly invited political inputs. While there is much to be said for explicitness and openness in a democratic society, the pros and cons of systematic analysis in politics are too complex to be explored here. In closing, it should be noted that citizens and politicians do not have a universal, consistent, and strong preference for explicitness and clarification of the sources of winners and losers. The failure of applied CBA to reflect the evolution of second-best economic theory is only partly due to reluctance of economists to relinquish the role of supplying authoritative advice.[7] The other part is a substantial public demand for self-deception and vainglory where we advertise a concern for the poor, human life, environment, or whatever, while acting selectively to the contrary. When politicians stand aside from resolving the conflicts of interest behind the CBA rules, they are able to embrace its results piecemeal—accepting its added legitimacy when it suits them, but labeling it academic irrelevance when they reject the results. If politicians were part of the systems analysis process, they would have to change its rules rather than selectively reject its conclusions.

NOTES

1. For a formal mathematical review, see (Dreze and Stern, 1987).
2. C.i.f. = import cost, insurance, and freight to the entry port.
3. For a review of contingent valuation use in the European Union, see (Bonnieux and Rainelli, 1999).
4. Sugden and Williams (1978, pp. 127-131) try to finesse the issues by assuming zero income effects. Reasonable applied analysts seem to differ on the reality of this assumption.
5. The Kaldor compensation test takes the *status quo* as the reference point and the winners compensates the losers. Losers are sellers. The Hicks compensation test requires the losers to pay potential winners not to undertake the project. Losers are buyers. Thus both actual and potential Pareto improvements have a starting place in property rights.
6. One problem with weighting incomes is that the resulting transfers and income changes are derivative. A citizen or politician can't know when deciding on a weight of 0.2 or 0.3 just how much income transfer is implied. The weight by itself is meaningless. A specific transfer budget with identified targets would be more explicit.
7. Other reasons for not pursuing systematic, explicit choice include (Wildavsky, 1969) argument that ambiguity is necessary to prevent political breakdown and ultimately civil war. (Leibenstein, 1987) makes a related point arguing that slack keeps firms with internal conflicts from coming apart.

REFERENCES

Adler, M. D. and Posner, E. A. (1999). *Rethinking Cost-Benefit Analysis* (Working Paper No. 72). Chicago: Law School, University of Chicago.

Akehurst, R. L. and Culyer, A. J. (1974). "On the Economic Surplus and the Value of Life." *Bulletin of Economic Research* 26: 63-78.

Arrow, K. J. (1963). *Social Choice and Individual Values.* New Haven: Yale University Press.

Arrow, K. J. and Fisher, A. (1974). "Environmental Preservation, Uncertainty, and Irreversibility." *Quarterly Journal of Economics* 88 (2): 312-319.

Arrow, K. J. and Lind, R. C. (1970). Uncertainty and the Evaluation of Public Investment Decisions. *American Economic Review* 60: 364-378.

Ball, M. (1979). "Cost-Benefit Analysis: A Critique." In *Issues in Political Economy: A Critical Approach,* Green, F. and Nore, P., eds. London: Macmillan.

Bateman, I. J. and Willis, K. G. (eds.) (1999). *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries.* Oxford: Oxford University Press.

Boadway, R. W. and Bruce, N. (1984). *Welfare Economics.* Oxford: Blackwell.

Bockstael, N. E. (1995). "Travel Cost Methods." In *The Handbook of Environmental Economics,* Bromley, D. W., ed. Cambridge MA: Blackwell.

Bockstael, N. E., Strand, I. E., and Hanemann, M. (1987). "Time and the Recreation Demand Model." *American Journal of Agricultural Economics* 69: 293-302.

Bonnieux, F. and Rainelli, P. (1999). "Contingent Valuation Methodology and the EU Institutional Framework." In *Valuing Environmental Preferences,* Bateman, I. J. and Willis, K. G., eds. Oxford: Oxford University Press.

Brown, G. M. and Shogren, J. F. (1998). "Economics of the Endangered Species Act." *Journal of Economic Perspectives* 12 (3): 3-20.

Campbell, D, T. and Stanley, J. C. (1963). *Experimental and Quasi-experimental Designs for Research.* Chicago: Rand-McNally.

Campen, J. T. (1986). *Benefit, Cost, and Beyond: The Political Economy of Benefit-Cost Analysis.* Cambridge: Ballinger.

Chipman, J. S. and Moore, J. C. (1980). "Compensating Variation, Consumers' Surplus and Welfare." *American Economic Review* 70: 933-949.

Cummings, R. G., Brookshire, D. S., and Schulze, W. D. (eds.) (1986). *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method.* Totowa NJ: Rowman & Allanheld.

Daly, H. E. (1991). "Towards an Environmental Macroeconomics." *Land Economics* 67 (2): 255-259.

Dasgupta, A. K. and Pearce, D. W. (1972). *Cost Benefit Analysis: Theory and Practice.* London: Macmillan.

DeAlessi, L. (1969). "Some Implications of Property Rights structures for Investment Choices within Government." *American Economic Review* 79: 13-24.

Dreze, J. and Stern, N. (1987). "The Theory of Cost-Benefit Analysis." In *Handbook of Public Economics,* Auerbach, A. J. and Feldstein, M., eds.. Vol. 2, pp. 909-990. Amsterdam: North-Holland.

Eckstein, O. (1961). "A Survey of the Theory of Public Expenditure Criteria." In *Public Finances: Needs, Sources, Utilization.* Princeton: National Bureau of Economic Research.

Freeman, A. M, (1979). *The Benefits of Environmental Improvement.* Baltimore: Johns Hopkins University Press.

Freeman, A. M. (1993). *The Measurement of Environmental and Resource Values.* Washington: Resources for the Future.

Gittinger, J. P. (1982). *Economic Analysis of Agricultural Projects.* Baltimore: Johns Hopkins University Press.

Gum, R. L. and Martin, W. E. (1975). "Problems and Solution in Estimating the Demand for and Value of Rural Outdoor Recreation." *American Journal of Agricultural Economics* 57: 558-566.

Hanemann, M. (1991). "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review* 81 (3): 653-647.

Hardin, E. and Borus, M. (1971). *The Economic Benefits and Costs of Retraining.* Lexington MA: Heath Lexington.

Harrison, A. J. and Quarmby, D. A. (1974). "The Value of Time." In *Cost-Benefit Analysis,* Layard, R., ed. Baltimore: Penguin.

Harrison, D. and Rubinfeld, D. L. (1978). "The Air Pollution and Property Value Debate: Some Empirical Evidence." *Review of Economics and Statistics* 60: 635-638.

Haveman, R. and Krutilla, J. V. (1968). *Unemployment, Idle Capacity, and the Evaluation of Public Expenditures.* Baltimore: Johns Hopkins Press.

Hoehn, J. P., Berger, M. C., and Blomquist, G. C. (1987). "A Hedonic Model of Interregional Wages, Rents, and Amenity Values." *Journal of Regional Science* 27: 605-620.

Hoehn, J. P. and Randall, A. (1989). "Too Many Proposals Pass the Benefit Cost *Test" American Economic Review* 79 (3): 544-551.

Jones-Lee, M. W. (1976). *The Value of Life: An Economic Analysis.* Chicago: University of Chicago Press.

Kelman, S. (1981). "Cost-Benefit Analysis: An Ethical Critique." *Regulation* 5 (January-February): 33-40.

Kopp, R. J. and Portney, P. R. (1999). "Mock Referenda for Intergenerational Decisionmaking." In *Discounting and Intergenerational Equity,* Portney, P. R. and Weyant, P. P., eds., pp. 87-98, Washington: Resources for the Future.

Krutilla, J. V. (1967). "Conservation Reconsidered." *American Economic Review* 57 (4): 777-786.

Kuhn, T. E. (1962). *Public Enterprise Economics and Transport Problems.* Berkeley: University of California Press.

Leibenstein, H. (1987). *Inside the Firm: The Inefficiencies of Hierarchy.* Cambridge: Harvard University Press.

Lesser, F. and Zerbe, Richard O., Jr. (1998). "A Practitioner's Guide to Benefit-Cost Analysis." In *Handbook of Public Finance,* Thompson, F. and Green, M. T, eds. New York: Marcel Dekker.

Lewis, A. (1972). "Reflections on Unlimited Labor." In *International Economic Development: Essays in Honor of Raul Prebisch,* Marco, L. D., ed. New York: Academic Press.

Little, I. M. D. (1957). *A Critique of Welfare Economics.* London: Oxford University Press.

Little, I. M. D. and Mirlees, J. A. (1974). *Project Appraisal and Planning for Developing Countries.* New York: Basic Books.

Majone, G. (1989). *Evidence, Argument and Persuasion in the Policy Process.* New Haven: Yale University Press.

McFadden, D. (1994). "Contingent Valuation and Social Choice." *American Journal of Agricultural Economics* 76 (Nov.): 686-708.

Mishan, E. J. (1976). *Cost-Benefit Analysis,* 2nd ed. New York: Praeger.

Mishan, E. J. (1981). *Economic Efficiency and Social Welfare.* London: Allen and Unwin.

Mishan, E. J. (1982). "The New Controversy about the Rationale of Economic Evaluation." *Journal of Economic Issues* 16: 29-47.

Mishan, E. J. (1988). *Cost-Benefit Analysis,* 4th ed. London: Unwin Hyman.

Pearce, D. W., Markandya, A., and Barbier, E. (1989). *Blueprint for a Green Economy.* London: Earth Scan.

Pearce, D. W. and Nash, C. A. (1981). *Social Appraisal of Projects: A Text in Cost-Benefit Analysis.* London: Macmillan.

Pohl, G. and Mihaljek, D. (1992). "Project Evaluation and Uncertainty in Practice: A Statistical Analysis of Rate-of-return Divergences of 1,015 World Bank Projects." *World Bank Economic Review* 6 (2): 255-277.

Robison, L. J. and Barry, P. J. (1996). *Present Value Models and Investment Analysis.* Northport AL: Academic Page.

Samuels, W. J. and Schmid, A. A. (1997). "The Concept of Cost in Economics." In *The Economy As A Process of Valuation,* Samuels, W. J., Medema, S. G., and Schmid, A. A., eds. Cheltenham: Edward Elgar.

Samuelson, P. (1963). *Foundations of Economic Analysis.* Cambridge: Harvard University Press.

Schmid, A. A. (1989). *Benefit-Cost Analysis: A Political Economy Approach.* Boulder: Westview Press.

Sen, A. (1979). "The Welfare Basis of Real Income Comparisons." *Journal of Economic Literature* 17(1): 1-45.

Sen, A. (1985). *Commodities and Capabilities.* Amsterdam: North Holland.

Shabman, L. and Stehenson, K. (1996). "Searching for the Correct Benefit Estimate: Empirical Evidence for an Alternative Perspective." *Land Economics* 72 (4): 433-449.

Slesnick, D. T. (1998). "Empirical Approaches to the Measurement of Welfare." *Journal of Economic Literature* 36 (Dec.): 2108-2165.

Smith, R. J, (1971). "The Evaluation of Recreation Benefits: The Clawson Method in Practice." *Urban Studies* 8 (2): 82-102.

Solo, R. M. (1999). "Foreword." In *Discounting and Intergenerational Equity,* Portney, P. R. and Weyant, J. P., eds. Washington: Resources for the Future.

Stiglitz, J. E. (1987a). "The Causes and Consequences of the Dependence of Quality on Price." *Journal of Economic Literature* 25: 1-48.

Stiglitz, J. E. (1987b). "The Wage-Productivity hypothesis: Its Economic and policy Implications." In *Modern Developments in Public Finance,* Boskin, M., ed. Oxford: Basil Blackwell.

Stilwell, F. (1999). "Cost-Benefit Analysis." In *Encyclopedia of Political Economy,* O'Hara, P. A., ed. London: Rutledge.

Sugden, R. and Williams, A. (1978). *The Principles of Practical Cost-Benefit Analysis.* Oxford: Oxford University Press.

Tengs, T. and Graham, J. D. (1996). "The Opportunity Cost of Haphazard Social Investments in Life Savings." In *Risks, Costs and Lives Saved,* Hahn, ed. Oxford: Oxford University Press.

Tresch, R. W. (1981). *Public Finance.* Plano TX: Business Publications.

Vatn, A. and Bromley, D. (1994). "Choices without Prices without Apologies." *Journal of Environmental Economics and Management* 26: 129-148.

Vining, A. R. and Wimer, D. L. (1998). "Passive Use Benefits: Existence, Option, and Quasi-Option Value." In *Handbook of Public Finance,* Thompson, F. and Green, M. T., eds. New York: Marcel Dekker.

Viscusi, W. K. (1996). "Economic Foundations of the Current Regulatory Reform Efforts." *Journal of Economic Perspectives* 10: 119-131.

Wildavsky, A. (1969). "Rescuing Policy Analysis from PPBS." *Public Administration Review* 29: 198-220.

Zerbe, R. O. and Dively, D. D. (1994). *Benefit-Cost Analysis In Theory and Practice.* New York: HarperCollins.

Chapter 16

LOCAL PUBLIC FINANCE

Charles B. Blankart
Humboldt-Universität zu Berlin and Rainald Borck
DIW Berlin
blankart@wiwi.hu-berlin.de

Rainald Borck
Deutsches Institute für Wirtschaftsforschung
rborck@diw.de

Abstract   In this essay, we survey the literature on local public finance. The first part deals with the normative theory of local public finance, starting from the question when it is beneficial to decentralize public services. We then analyze the functioning of a system of competitive jurisdictions in the spirit of Tiebout. The final part of the essay deals with constitutional design. In particular, we ask when and how local governments have to be regulated in order to prevent destructive competition or contain monopoly power, and we describe which institutions might perform these tasks.

Keywords: Local public finance, decentralization, Tiebout model

JEL classification: H7

Local public finance has a long tradition which may be traced back at least to the town founding era in the high middle ages (900–1350). In contrast to ancient towns and cities which had existed for many centuries as capitals and cultural or commercial centers, the towns established in the high middle ages can be regarded as entrepreneurial investments by sovereigns and other regional power holders who aimed at extending their domain. The German economic historian Dirlmeier concludes: "In the same way as a hostile fortress is neutralized by constructing an own fortress in close neighborhood, market places and towns are established and privileged in pursuit of immediate competition" (Dirlmeier, 1966, p. 3, translated by the authors). Sovereigns and

power holders granted their towns local autonomy, including in particular the right to establish their own legal order with equality before the law for the different classes of citizens living within the town walls, in short, a climate favorable for market exchange (Ennen, 1972, p. 11). In return, sovereigns and power holders claimed taxes, and other charges, such as returns from the mint. Hence, local public finance in these early times was less a problem of how the town paid for their own public services, but of how a sovereign or power holder could extract the return on his investment.

Usually the power holders claimed local taxes as a lump sum which a town was liable to pay per period, but which it could collect according to its own rules. These terms of payment turned out to be relatively conducive to urban economic development. For the towns had an incentive to raise taxes according to general non-distorting criteria. Moreover, the obligation to pay taxes in money, rather than in labor services (as under traditional feudalism), promoted the use of market exchange. This form of taxation can be regarded as the general rule from which exceptions, sometimes considerable ones, were made when wars or large infrastructure projects had to be financed.

This rule was given up in continental Europe towards the end of the 17th century. Under the influence of French absolutism, sovereigns increasingly decided to send their own tax collectors into the towns and to assess and differentiate taxes on the spot. Moreover, excise taxes levied at the town gates became more and more important. Although this system generated higher revenues in the beginning, it became hated and inefficient in the longer run because its differentiation was regarded as arbitrary, and because it discouraged investments which were the base for future economic growth. On the eve of the French revolution, the average tax burden per capita was only about half as high in France as it was in the United Kingdom; however, whereas this relatively low tax burden in France was combined with fiscal exploitation on the one hand and tax privileges on the other, in the United Kingdom—where arbitrary tax differentiation had been prohibited since the Petition of Rights (1628) and the Bill of Rights (1689)—the relatively high tax burden was more conducive to economic growth. Hartmann (1978) argues that this inequity and inefficiency of the French tax system contributed considerably to the outbreak of the revolution in France in 1789, whereas civil loyalty prevailed in the United Kingdom.

During the revolution the traditional tax system in France was swept away and replaced by a system which relied heavily on the taxation of land and local crafts. Insofar, it approached the British system which traditionally relied heavily on local land taxation. The British system also served as a model for the U.S. system of local property taxation.

Today, local governments in most countries still rely on these traditional taxes. Yet local tax bases have become too small to support higher level gov-

ernments (formerly the sovereigns) and too small to even finance their own needs and the tasks imposed upon them by upper level governments. Local governments, therefore, heavily depend on grants and transfers from upper level governments. Moreover, the tax autonomy of local governments has been restricted. Only in few countries such as the United States, Canada, Australia and Switzerland do local governments still enjoy substantial tax autonomy. Germany takes a middle position. Its local governments may tax land and local business (within federal regulations), but they cannot tax physical persons and insofar have to rely on grants and tax sharing.

While local public finance has a long history, the records of the theory of local public finance are much shorter. The 17th and 18th century cameralists had a pragmatic view of the state. They saw the whole economy and thus also the local communities as a resource serving to maximize their sovereign's wealth. Therefore, they asked what a good fiscal policy should look like to achieve this goal. But they did not regard localities as self organized local communities (see e.g., Tautscher, 1947 for the cameralist view). Such aspects began to emerge only in the 1930s. Johannes Popitz, an academic of Berlin University, secretary of state and later minister of public finance in the state of Prussia, felt that local governments would become increasingly unable to fulfill their local tasks because of a lack of tax resources and that they would increasingly lose their autonomy in favor of upper level governments. He believed in the "attractive power of the highest-level purse" (Popitz, 1927, p. 348, translated by the authors). At about the same time, the German economist Arnold Brecht found in statistical records that larger cities have larger per capita expenditures because of diseconomies of scale and because of the larger bundles of services they provide (Brecht, 1932).

Letting aside these forerunners, the effective start of the theory of local public finance has to be placed in the second half of the 20th century with two seminal articles: Charles Tiebout's (1956) paper "A Pure Theory of Local Expenditures," and Wallace E. Oates' (1968) article "The Theory of Public Finance in a Federal System." These two papers still form the basis of the current scholarly discussion. They will serve as the backbones of our survey. Interestingly, Oates' approach, though posterior, attracted more attention initially. It focuses mainly on what governments, in particular local governments, ought to do and hence fits within the normative theory of public finance which was the dominant view following Musgrave's seminal book *The Theory of Public Finance* (1959). Therefore, we shall first review this normative approach in the following Section 1. The conceptual viewpoint taken by these models is to ask: What would an optimal system of local (and, if necessary, higher level) governments look like, if one could draw up such a system from scratch? In contrast, positive theories of local government ask: How does a system of local governments function under certain assumptions? At the heart of our treatment

in Section 2 will be the Tiebout model: It explains local government performance under perfect mobility of citizens. Tiebout showed that under restrictive assumptions, local governments supply public services efficiently. This point of view has become particularly popular since migration of labor and even more of capital has steadily increased with the integration of regional and national economies. The section also contains normative aspects, namely, various failures of the Tiebout mechanism to achieve an efficient allocation of resources. If the Tiebout mechanism does not function perfectly, from a normative point of view one needs to analyze alternative mechanisms to ensure an efficient supply of public goods. In Section 3, we discuss some alternatives based on the constitutional or public choice approach. Its idea is to analyze alternative collective decision making procedures and to make proposals for a reform of local governments' constitutions. We will argue that the need for such an approach follows from the fact that the Tiebout mechanism cannot be expected to function perfectly. Therefore, local politics matter. In the concluding section, we briefly outline what we believe are fruitful areas for future research in local public finance.

1.   THE NORMATIVE THEORY OF LOCAL PUBLIC FINANCE

1.1.   The Conceptual Point of View

Suppose one had to design a federal system from scratch, what would the proper tasks of local governments be? That is the conceptual viewpoint taken by the normative theory of local public finance. Following Musgrave's (1959) triad, Oates (1972) investigated the proper functions of local governments in stabilization, redistribution, and allocation. He argued that local activities to stabilize the economy would be ineffective because the incidence of expenditure programs and tax cuts would spill over into neighboring jurisdictions and hence create benefits outside the constituency who pays for them in the form of larger public debt. Similarly, local redistribution programs would fail. Rich people would have an incentive to outmigrate and poor people immigrate if one jurisdiction started a redistribution program in isolation.

Concerning allocation, however, matters are said to be different. The incidence of many public services such as schools, hospitals, police, waste removal, parks and recreation, water, sewage, cable grids etc. is mainly local and provision should therefore fall into the responsibility of local governments. Local decision and provision are indicated because very often preferences differ among localities. Purely residential local communities have different needs than communities with large factories or communities which provide rail, motorway or air transport facilities and those with touristic amenities.

1.2.   Optimal City Size

To determine what local governments should do is one question. Another is in what city size these governmental functions should be fulfilled. In reality we can find very large but also very small local governments. New York City has about 10 million inhabitants, while small towns may have only a few hundred. There has been a lively debate in the seventies on which city size is most efficient. Central government administrations have often argued for larger cities as being more efficient, whereas local citizens have opposed such interferences into their community life.

The discussion in economics was opened again by Oates (1972) who made the case for small local governments. His decentralization theorem

> indicates that in the absence of cost-savings from the centralized provision of a [local public] good and of interjurisdictional externalities, the level of welfare will always be at least as high (and typically higher) if Pareto-efficient levels of consumption are provided in each jurisdiction than if any single, uniform level of consumption is maintained across all jurisdictions (Oates, 1972, p. 54).

Although in principle, central governments could replicate any allocation achieved by local governments, the literature usually assumes that for political or informational reasons, the central government is confined to providing uniform levels of public services to citizens regardless of their place of residence. For instance, in some countries, central governments are prohibited by the constitution to discriminate public service levels by location. The idea that local government officials have better information on local preferences and, therefore, are better able to tailor outputs to local tastes, has also been widely accepted.

While Oates focused mainly on local versus higher level governments, one can also conclude that, among local governments, small units can better adjust to varying preferences than larger local governments, or conversely that large local governments are less able to provide differentiated public outputs to the various groups within a local jurisdiction because of transactions and information costs. Consequently, there are welfare gains in providing public goods in small local jurisdictions, ceteris paribus. The welfare gains are larger, the less price elastic the demand for pubic goods, and they may be large in practice because demand for publicly provided goods is typically inelastic (Oates, 1999, p. 112).

Another argument for smallness comes from search and experimentation. A large number of small governments allows for more variety than a few large governments. More alternative forms of local government organization can be tested, and at costs which are smaller than with few large governments. This is reflected in the popular phrase of state and local governments as "laboratories."

A more indirect benefit of small jurisdictions is that political participation is likely to be higher. If this is positively valued, smallness carries an addi-

tional benefit besides the traditional 'economic' benefits (Inman and Rubin-feld, 1997a).[1]

While small local governments allow for supply adjusted to diverse preferences, economies of scale from joint consumption of public services set a limit to an ever decreasing size of local governments. Hence, there exists a trade-off between smallness and scale economies. This is, however, only half of the truth. In an early paper Tullock (1969) argued that the scale problem can be overcome by outside contracting. Small local communities could contract with large private suppliers or other governments to obtain a service. Therefore, there should be no need for small communities to merge in order to exploit these economies. In fact, Fisher (1996) cites reports that in 1982, about half of all city and county governments in the U.S. contracted with other governments to provide some services. Contracting therefore seems to be widely used as a method to exploit economies of scale.

Tullock's argument is, however, not fully convincing either, since not all services can be contracted for easily. Some are hardly tangible and some exhibit economies of scale with sunk costs. For the former services a contract cannot be easily stipulated and monitored. Therefore, it is not unambiguously clear what the terms of contract are. These difficulties may lead to adverse selection, moral hazard, and to a deterioration of the service at high budgetary costs. Moreover, the supplier may behave strategically because he knows that the government cannot easily switch to another provider. Under such circumstances it may be preferable for local governments to make rather than buy these services, and the well-known trade-off between preference adjustment and scale economies re-emerges. Hardly tangible services such as social assistance for poor and disabled persons may be carried out more carefully by the local government than by a private contractor who substitutes profits for unobservable quality characteristics. Similarly, a local government may prefer to operate its own sunk costs facilities such as fresh water grids and sewage in order to be independent of the vagaries and the threat potential of a private supplier. In practice, contracting is widespread for intermediate inputs, but also for solid waste collection and disposal and transportation. In other categories such as public safety and health services, it may be hard to specify exactly what the desired service is, and, therefore, contracting is much less prevalent (Pack, 1991).

Still another way for local governments to escape the trade-off between preference adjustment and scale economies is to establish special districts. Several jurisdictions, each being too small to provide the service on its own, join to set up a special organization to provide the service for all of them. Special purpose units are common for the sunk costs facilities mentioned above (fresh water and sewage) as well as for waste disposal facilities, airports, schools, theaters, etc. Often a group of local governments runs a re-

gional public transit system as a special purpose unit. Routes and time tables are harmonized and a common ticket system applied to all suppliers. Revenues are distributed according to a predetermined key. Although such special districts have been welcomed by many users, they embody problems of their own. Their management is often relatively independent from local governments and far away from voters' control. Moreover, market entry by outsiders is often made difficult—even if actual or potential competition were feasible—because it may reduce common profits or increase common losses for which all participants have to pay. Finally, federal or state subsidies (if any) have to be shared among more suppliers. Therefore, special purpose units may tend to become large uncontrolled bureaucracies, inimical to competition, whose scale economies tend to be eaten up by behavioral inefficiencies.

Recently Frey and Eichenberger (1999) have proposed to overcome some of the deficiencies of special districts by organizing local and regional governments by function rather than spatially. Instead of local governments they propose "functional, overlapping, competing jurisdictions," so-called "FOCJ." Supply should not come from traditional multiproduct local governments, but from these functionally specialized jurisdictions, competing with each other. Customers should shop around and demand services independently of their place of residence. They may contract collectively through their local community or individually. Pricing is organized collectively by FOCJ's members, their customers. Since FOCJ are given taxing power, they can charge their customers differently. It is not quite clear, however, what services could be supplied along this concept characterized by free choice and collective pricing. In non-increasing returns industries, individuals will switch to other suppliers if the prices assigned to them are above costs. Competition would lead to marginal cost pricing for each publicly supplied good and (since consumers "sort" into homogeneous FOCJ) collective choice would be immaterial. By contrast, in increasing returns industries, collective decisions on charging different prices may be enforceable because competition is generally prevented. Hence an institutional choice has to be made between either competition with one price or regulated monopolies with collective price setting and price discrimination. It is difficult to see how both competition and collective choice together could be achieved in a meaningful way.

The problem of optimal city size has also been approached by urban economists (see Mieszkowski, 1987). They have studied optimal city size as it emerges from a trade-off between economies and diseconomies of agglomeration. On the one hand, there are economies of scale from joint consumption of public goods or from increasing returns to scale technologies. Furthermore, there are production economies, which can be of two kinds: localization economies, which result from increasing returns in the output of an

industry (e.g., sharing of inputs, communication economies), and urbanization economies which depend on the size of the metropolitan area. Since the former seem to dominate the latter, it has been concluded that cities should specialize in particular industries. In some models it is shown that cities may be too small. Thus, Tolley (1974) argues that congestion and pollution externalities reduce residents' real income below its optimum and limit city size to a suboptimal level. Therefore, in determining optimal city size, not only preference diversity, but also scale, congestion and production economies have to be considered.

1.3.   Spillover Effects

Oates (1972) suggested that local governments should provide public goods when there are no interjurisdictional externalities. While pure local public goods are defined such that the benefits of the goods accrue only to the residents living within a jurisdiction, many public goods do not have this characteristic. For instance, the external benefits of an educated citizenry are likely to materialize largely outside of the jurisdiction providing public education. Pollution is the prime example of a negative spillover. Olson (1969) has proposed to organize jurisdictions according to the principle of fiscal equivalence. Benefits of local public services should accrue to those who pay the taxes to finance them, and taxes should be borne only by those who enjoy the benefits.[2] Where this principle is violated, local governments will generate spillover benefits or spillover costs (alternatively: external benefits or external costs). Since local governments, taking into account their residents' welfare only, will ignore spillovers, the supply of a public service generating spillover benefits will be too small and that of a public activity producing spillover costs too large.

The problem of spillover benefits has been discussed within the central city exploitation thesis and the corresponding hypothesis of the flight into the suburbs (Neenan, 1970). It has been argued that the suburbians use the facilities of the central city without paying taxes for their provision. Conversely, the central cities are said to pollute the environment of the suburbs through their industrial activities.

A straightforward way to overcome these problems could be the application of the Coase Theorem. Given well defined property rights, local governments will have an incentive to internalize spillover benefits and spillover costs through bargaining.[3] Often the internalization of spillover benefits is easier than that of spillover costs because the former can be localized at particular facilities such as theaters, schools, universities, hospitals etc., whereas the latter are dispersed in the form of general multi-source pollution, involving many instead of a few negotiating parties. For instance, the city of Zurich is successful in periodically making arrangements with its surrounding local communities on cost sharing of cultural and other city services. Similarly, the city

of Basel has reached an agreement on sharing the costs of the university with the surrounding canton of Basel-Land and the other cantons. In both cases, the arrangement was facilitated because institutions of co-decision making on the levels and qualities of the services were set up (see Pommerehne and Krebs, 1991 for the case of Zurich). Conversely, environmental problems are rarely handled among local communities because the costs of negotiating among a large number of participants are too high, though there are exceptions such as the Ruhr valley waste water regulation. In general, however, central govern-ments tend to impose the necessary environmental regulations.[4]

An alternative way to approach the urban spillover problem is by consoli-dating local governments (Hoxby, 1996; Gilbert and Picard, 1996). The trade-off to be solved is that on the one hand a consolidated local government will suffer less from spillovers, but on the other it will lack the differentiated infor-mation on preferences and production costs to accomplish an efficient alloca-tion.

1.4.   Intergovernmental Grants

Another straightforward way to deal with spillovers is through intergovern-mental grants.[5] To internalize positive spillover effects, the central government should give the local governments open-ended matching grants: for each dollar spent on a public good whose benefits spill over into other jurisdictions, a local government would receive a specified matching rate, which would correspond to the marginal spillover benefits received by other jurisdictions.

The normative theory of local finance has also recommended unconditional grants, mainly with the goal of achieving redistributional goals or some notion of horizontal or vertical equity.[6] With respect to redistribution, it is argued that use of progressive taxes by local governments for redistribution is likely to fail; therefore, the central government should retain the redistributive function (see Section 1.2). It would then effectively act as a revenue collecting agent for local governments and then redistribute the collected revenue through in-tergovernmental grants.

Moreover, unconditional grants may be used to ensure horizontal equity. For instance, with proportional (or progressive) income taxes, an individual would pay lower taxes for the same level of public services if she lived in a rich community than if she lived in a poor community. Likewise, with increasing returns from joint consumption of public goods, lower tax levels would result in more populated jurisdictions. To restore horizontal equity in these cases would require transfer payments.[7]

An important topic in local public finance is the analysis of local govern-ments' response to unconditional grants. Bradford and Oates (1971) showed that if the supply of public goods is determined by simple majority vote, public

spending should react to an increase in unconditional grant money in the same way as it would react to an equivalent increase in private income. However, empirical studies have found time and again that the response of spending to unconditional grants is much higher than the response to increases in income. This has been dubbed the flypaper effect, or "money sticks where it hits."

Essentially, the response to this 'anomaly' has been of three types:[8] one argues for misspecification; another tries to argue that the flypaper effect is really consistent with rational decision making by voters; and the third type explains the existence of the flypaper effect in models which depart from the median voter model. Studies of the first type argue, e.g., that some grants are treated as unconditional which are really not. The second type of studies has employed different arguments: for instance, if local taxes are distortionary, then the price of public goods is lowered by unconditional grants which are not distortionary; a larger increase of spending in response to grant money than in response to income increases is then rational (Hamilton, 1986). More-over, if the median voter's local tax share is larger than his federal tax share, a one Euro increase in grants enriches the median voter by more than a Euro (Fisher, 1979).[9] The third type assumes that local decisions are taken by bud-get maximizing bureaucrats. If voters are not fully informed about grants or the marginal tax price of public goods, bureaucrats can use grants to increase spending above the level desired by voters (Oates, 1979; Filimon, Romer and Rosenthal, 1982). A necessary condition for bureaucrats to be able to exploit voters is the existence of imperfect information or fiscal illusion. Hines and Thaler (1995) argue that the flypaper effect stems from individuals treating money on hand (grants) differently than money which has to be raised through taxation. It appears from the literature that flypaper effects are persistent to more careful empirical specifications, which points to the need of explicitly modelling local government decision making (or, if the type of anomaly sug-gested by Hines and Thaler (1995) is prevalent, to a "behavioral theory of local finance").

2.   THE POSITIVE THEORY OF LOCAL GOVERNMENTS

2.1.   The Tiebout Model

Without doubt Tiebout's (1956) paper is the most influential article on local public finance. Tiebout advanced the important idea that local governments would provide public goods to mobile consumers in an efficient manner. The idea was strikingly simple. Local governments attract citizens through pro-viding tax/public good packages until achieving an optimum community size. By choosing their place of residence, consumers choose the tax/public goods

package that best suits their preferences. Therefore, in equilibrium, no consumer can be made better off by moving, and the equilibrium is Pareto efficient. Tiebout offered local public goods provision as a solution to the problem posed by Samuelson (1954) that consumers would not reveal their preferences for public goods, and therefore an efficient supply of public goods will most likely fail in practice. While Tiebout noted that various imperfections will limit the efficiency of local public good supply in practice, he concluded by writing that those who compare the reality of local government with the reality of competitive markets *"may* find that local government represents a sector where the allocation of public goods [...] need not take a back seat to the private sector" (Tiebout, 1956, p, 424, emphasis in original).

In this section, we will describe Tiebout's original assumptions and show how they lead to an efficient allocation. The next subsections then look at problems that come up when the assumptions are relaxed. In particular, Tiebout assumed the following.

1. Consumers are completely mobile and move to the jurisdiction that best satisfies their preferences.
2. Consumers are fully informed about communities' taxes and public goods supply and move in response to differences in taxes or spending levels.
3. There is a large number of communities among which consumers may choose.[10]
4. Individuals live on dividend income, so there is no requirement that individuals live and work in the same jurisdiction.
5. Governments use lump-sum taxes or user charges to finance public services.
6. There are no interjurisdictional spillovers, that is, public goods are purely local public goods.
7. The average cost per person of providing public goods to the population is a U-shaped function of the number of residents.
8. Communities attract residents exactly until the optimal city size—i.e., the minimum per capita cost—is reached.

To illustrate the workings of Tiebout type models, we begin with a simple model based on McGuire (1974). Essentially, local governments are treated like "clubs" (Buchanan, 1965).[11] To determine what optimal jurisdictions look like, suppose first that the economy is composed of N identical individuals, each with preferences characterized by a well-behaved utility function $U(X, G)$, where X is a private good and G a local public good. Individuals have income y which they can allocate between public and private goods. The costs of public goods in terms of the private good are given by a cost function $C(n, G)$, where n denotes the number of consumers in a jurisdiction. Suppose

the costs of the public good are to be shared equally by all individuals. Then the problem of determining optimal jurisdiction size can be stated as that of maximizing the utility of a representative citizen:

$$\max_{n,G} U\left(y - \frac{C(n,G)}{n}, G\right).$$

The first order conditions can be written as (subscripts denote partial derivatives):

$$\frac{C(n,G)}{n} = C_n, \tag{1}$$

$$n\frac{U_G}{U_X} = C_G. \tag{2}$$

A Pareto optimal allocation occurs when both of these conditions are fulfilled simultaneously. Equation (2) is the Samuelson condition for the efficient supply of public goods, according to which provision should be carried out until the sum of the marginal rates of substitution equals the marginal cost of the public good. Equation (1) is the membership condition, and states that each jurisdiction should admit individuals until the per capita cost is minimized, which occurs when the average cost per person equals the marginal cost of admitting an additional individual. Clearly, when average costs are decreasing at all levels of G (as with purely public goods), the optimal community would include the entire population, while increasing average costs would result in an optimal club size of zero.[12] With constant average and marginal costs, the optimal community size is indeterminate. Let $n*$ be the optimal jurisdiction size. Then the optimal number of jurisdictions is $N/n*$, assuming this is an integer.

Suppose instead that the population is heterogeneous, say, because individuals differ by income or preferences. Then the population should be partitioned into homogeneous clubs, each of size $n*$ (assuming that the cost function is such that marginal crowding costs are independent of the level of the public good). Heterogeneous clubs cannot be efficient in this model. If some jurisdiction were to contain both high and low demanders who consume the same amount of public goods, both types could be made better off by segregating into homogeneous communities with their own type (assuming, again, that there are enough individuals of each type to attain the optimal community size).[13]

Tiebout's claim was that local public goods provision results in an allocation which satisfies (1) and (2). To show how this can be accomplished, consider a "club good" model with profit maximizing clubs and free entry

(Buchanan, 1965; Cornes and Sandler, 1996). In the local public goods model, this would correspond to the assumption that jurisdictions can be formed freely on a featureless plain. Each jurisdiction chooses a head tax or access charge, T, to maximize profits (i.e., tax revenue minus costs). Since individuals are perfectly mobile, in each jurisdiction they must attain the same utility level, say, $\bar{u}$. We assume that jurisdictions treat this as exogenously given. This latter assumption is referred to as utility taking, and mimics the role of price taking in perfect competition. The jurisdiction's problem is then

$$\max_{T,G,n} \quad nT - C(n, G) \quad \text{s.t. } U(y - T, G) \geq \bar{u}.$$

Solving the maximization problem, we get the first order conditions

$$n\frac{U_G}{U_X} = C_G, \tag{3}$$
$$T - C_n = 0. \tag{4}$$

Equation (3) implies that local public goods will be supplied according to the Samuelson rule. (4) implies that individuals will be charged a head tax that exactly equals their marginal crowding costs. Together with the zero profit assumption, this implies that average and marginal crowding costs will be equated. Thus, the equilibrium with profit maximizing clubs will produce an efficient allocation, both with respect to the supply of public goods within, and the allocation of consumers among communities.

In a more rigorous paper, Bewley (1981) showed that an equilibrium exists and is efficient if communities maximize profits and per capita costs are constant. However, Bewley showed that the assumptions needed to prove optimality of Tiebout equilibria are very restrictive. In particular, by requiring governments to maximize profits and provide quasi-private goods, the Tiebout problem is essentially the same as that of existence and optimality of equilibria in competitive private goods markets.

In the following subsections, we consider what happens when some of the assumptions of the Tiebout model are relaxed. Before doing so, however, one should note that the model we have presented here as well as that in Bewley (1981) lack one distinguishing feature of local public goods economies, namely, the role of space. In order to consume local public goods, consumers must reside in the jurisdiction where they are provided. Space thus constitutes an essential component of local public goods models. For now, we note that it is possible to amend the Tiebout model with space without changing the results. Scotchmer (1994) and Wildasin (1986), among others, have shown that in an economy with local public goods, if jurisdictions are managed by

"developers" who maximize property values, a Tiebout equilibrium will be efficient. The reason for this result is that developers realize that changing taxes or public goods supply will affect the value of property in their community. Thus, capitalization of local fiscal policies leads property value maximizing developers to provide public services efficiently.

2.2.   Critique of the Tiebout Model

For reasons of space, we treat only a small subset of the consequences of relaxing the assumptions of the Tiebout model. In particular, we will focus on three of them: increasing returns to scale, distortionary taxation, and restrictions of choice due to immobility or a limited number of communities. The problem of spillovers has already been discussed in Section 1.3 and will therefore not receive any further attention.

2.2.1.   Increasing Returns to Scale   Increasing returns to scale might upset the efficiency properties of the Tiebout model. The literature usually analyzes economies of scale that do not stem from the technology of production, but rather from joint consumption of publicly provided goods. For pure public goods, the cost of providing any level of the good, G, is independent of the population, since "each individual's consumption of such a good leads to no subtraction from any other individual's consumption of that good" (Samuelson, 1954, p. 387). Hence one can write the cost function as $C(n, G) = C(G)$, where n is the number of people who consume the amount G.

Bewley (1981) presents an example where decentralized provision leads to an inefficient equilibrium. In his example, two jurisdictions provide identical levels of the public good and identical tax rates. This is an equilibrium since consumers have no incentive to move. All individuals could be made better off, though, if they lived together in one community. It appears though, that the problem in this example is that governments do not initiate policy changes to attract consumers.[14]

But other models also lead to inefficiency with pure public goods. Consider the "club model" presented in Section 2.1. Suppose that individuals are heterogeneous, i.e., some have a high and others a low willingness-to-pay for public goods. From the membership condition, it is clear that with purely public goods, the optimal membership consists of the entire population. But this clearly conflicts with the competitive analogy. Profit maximization with free entry would at best lead to an equilibrium with head taxes equal to average costs (assuming that perfect price discrimination or some other efficient mechanism is not feasible). As long as some individuals have a willingness-to-pay below this average-cost level but larger than zero, the allocation is not efficient. Thus, Pareto optimality would seem to be hard to attain with pure public goods.

It should be noted, however, that this result was derived in a model without spatially fixed factors. That is, one of the defining properties of local public finance models was ignored, namely, consumers have to occupy space (or housing) in order to consume public goods. If there are spatially fixed resources, crowding occurs even with purely public goods due to the scarcity of the fixed factor. Therefore, the optimal size of jurisdictions might be relatively small even with purely public goods. Scotchmer (1994) and Wildasin (1986) show that capitalization of public goods and taxes into property values leads local property value maximizing developers to supply an efficient level of public goods. But, in Scotchmer's (1994) model, this result is derived with an exogenously fixed number of jurisdictions. This may or may not correspond to the optimal number of jurisdictions, and, therefore, the allocation is only efficient contingent on the existing number of jurisdictions. However, it might be preferable to redraw jurisdictional boundaries.[15]

2.2.2.   Tax Competition with Distortionay Taxation   Problems arise in the Tiebout model if public services are financed by distortionary taxation. Since the literature has concentrated on property taxes, we will limit the discussion here to that particular method of financing expenditures, although similar arguments would apply to taxes on labor income, excise taxes, or other distortionary taxes. One problem that property tax financing may cause is nonexistence of an equilibrium (see, e.g., Rubinfeld, 1987). Suppose there are only two types of individuals, rich and poor, who live in segregated communities. Then each poor individual, assuming the level of public goods and taxes is unaffected by his migration decision, might have an incentive to move to the rich community, buy a small house, and consume more public goods at a relatively low tax price. Thus, the problem of "the poor chasing the rich" may threaten existence of an equilibrium. But existence can be restored if zoning ordinances can dictate the choice of a minimum lot size (Hamilton, 1975). In this case, poor people would be prevented from migrating, because buying a big house would be unattractive for them. Epple, Filimon and Romer (1984) discuss sufficient conditions to ensure existence of equilibrium with property taxes. In their case, however, the number of communities is fixed so equilibria will generally not be efficient (see Section 2.2.4).

The bulk of the literature on tax competition with distortionary taxes has focused on the possibility of underprovision of local public goods.[16] Put simply, if local governments use a source tax on mobile capital, tax competition will lead local governments to set inefficiently low tax levels for fear of driving out mobile capital. To make the point, consider the following model (Zodrow and Mieszkowski, 1986). Suppose the economy is composed of N identical jurisdictions, each endowed with a fixed factor, say, land. There is also a fixed national capital stock $\bar{K}$. Capital is assumed to be perfectly mobile, so it has

to earn the same net rate of return r in each jurisdiction. Capital is combined with the fixed factor to produce output according to the production function $F(K), F' > 0, F'' < 0,$ where K is the amount of capital employed in jurisdiction i.[17]

Citizens are identical and receive income in the form of land rent and capital income. Capital mobility implies that from the view point of each region, capital is a function of the gross-of-tax rate of return, $K(r + T)$, with

$$K'(r + T) = \frac{1}{F''} < 0.$$

Governments use tax revenue to cover the costs of a publicly provided private consumption good, G. It is assumed that each government maximizes the utility of a representative citizen, which is a function $U(X, G)$ of publicly provided goods and private consumption X. The first order condition for a maximum is

$$\frac{U_G}{U_X} = \frac{1}{1 + \tau\epsilon}, \tag{5}$$

where $\epsilon < 0$ is the elasticity of capital demand with respect to the gross return and $\tau := \frac{T}{r+T}$ is the ad valorem tax rate. Since the Samuelson rule would call for public good provision up to the point where the marginal rate of substitution equals the marginal rate of transformation (one), equation (5) implies that at the Nash equilibrium, public goods are underprovided.

The problem comes from the fact that increasing property tax rates involves a "fiscal externality." From the viewpoint of each region, the costs of increasing taxes consist of a revenue effect, (K), and a tax base effect $(T K')$. The latter term is negative, hence, the reduced incentives for using the capital tax. From the nation's view point, however, this is not a cost factor, since capital simply moves from one region to another. In fact, since the national capital stock is inelastic, a national tax on capital is not distortionary at all. The fiscal externality arises because each region ignores the positive effect that capital movements induced by tax increases produce in other regions. The resulting equilibrium is inefficient, and welfare could unambiguously be increased if all regions simultaneously increased their tax rates.

Thus, the literature on tax competition usually suggests some form of tax coordination to alleviate the fiscal externality problem. In the case of identical regions, it is easy to see that a coordinated increase of tax rates, starting from the Nash equilibrium, increases public good provision and welfare.

An alternative way to internalize the fiscal externality is for the central government to provide matching grants to the local jurisdictions. The matching rate can be chosen such that each jurisdiction provides public goods according to the Samuelson rule (Wildasin, 1989).

2.2.3.   Discussion and Extensions   While the formal models of the tax competition literature have yielded important insights, several criticisms have been put forward.[18] Most models of tax competition simply assume that the government must finance its expenditures by taxes on capital. If the government had access to a head tax, it would optimally set the property tax rate to zero and choose the level of the head tax that leads to a Pareto optimal supply of public goods (Zodrow and Mieszkowski, 1986; Oates and Schwab, 1988). As a result, the proposed methods for dealing with the fiscal externality problem are all second best approaches. McLure (1986) argues forcefully that the problem is not tax competition per se, but rather the use of inappropriate tax instruments. If governments had access to user charges or benefit taxes, the problem of tax competition would disappear (assuming there are no economies of scale and the other Tiebout assumptions hold). If this is not possible, then taxes should be levied on activities that most closely reflect the benefits of the public goods financed by the tax.[19]

One explanation for the use of capital taxes instead of more efficient labor taxes is the fact that these taxes redistribute income among heterogeneous voters. Consider a standard tax competition model where regions are large enough to affect the real return to capital. Suppose voters have differing endowments of capital and labor and that these are perfectly negatively correlated. The choice of capital and labor taxes is made by majority vote. Under suitable assumptions (namely, intermediate preferences), one can show that the voter with the median endowment within a jurisdiction is decisive. If this voter has a small capital endowment (large labor endowment), he may choose to tax capital instead of labor (Borck, 2003).[20] Using political economy arguments such as this one can rationalize the choice of capital taxes instead of simply assuming it.

But even distortionary taxes may have aspects of benefit taxation, if for instance, demand for the taxed goods correlates with demand for the public good. Another ingenious argument has been put forward by Hamilton (1975). He showed that if communities can use zoning to dictate the minimum value of property that a household must purchase to reside in the community, the property tax is converted into a pure benefit tax. The efficiency properties of the Tiebout model are then restored.[21]

Property taxation is the prime source of local finance in the U.S. (besides transfers from higher level governments) and hence, interest on this topic has been substantial. The real question then seems how efficient zoning measures are. Mieszkowski and Zodrow (1989) argue that for legal reasons, zoning measures are imperfect and hence, the property tax is not a benefit tax. For the contrasting view that despite legal limitations, communities do effectively achieve zoning, see Fischel (1992).

While earlier discussion on the desirability of tax competition has focused on the question of benefit taxation, more recent contributions have argued that there are reasons to believe tax competition to be (at least partially) efficiency enhancing. In essence, it is claimed that in some circumstances, government spending may be excessive, and tax competition and the implied downward pressure on spending may be a useful tool for constraining spending. One simple example of this is tax exporting. If a substantial part of local expenditures is financed by taxes which are borne by non-residents, local governments will have an incentive to increase spending beyond efficient levels. Examples of such taxes may be property taxes imposed on absentee landowners, sales or excise taxes on cross-border shopping, taxes on gambling, severance taxes, etc. McLure (1967) estimated that, on average, about 20 percent of state and local taxes were exported in 1962.

A second point is the interaction of horizontal and vertical tax competition, i.e., competition that arises from taxation of tax bases shared by different levels of government (see, e.g., Keen, 1998). Consider two distinct levels of government, say, federal and local, both of which levy taxes on the same tax base. Common examples are state or provincial sales and income taxes. Here, as in horizontal tax competition, fiscal externalities arise, but they are negative rather than positive: By raising its tax rate, one level decreases the tax base of the other. If each level ignores these externalities, equilibrium taxation will generally be too high.[22] If there is both horizontal and vertical tax competition, there are two opposing forces acting on tax rates, with the net effect being unclear.

Tax competition may also be desirable if it constrains Leviathan governments (Brennan and Buchanan, 1980). This is most easily seen by considering a Tiebout model, where governments maximize profits. If publicly provided goods are purely private, an equilibrium with competing governments exists and will be efficient (Bewley, 1981). However, if the publicly provided private good were supplied by a revenue maximizing central government, this would act as a monopolist and therefore raise too much taxes and supply too little of the public good. In this case, competition among governments serves the beneficial role of constraining Leviathan.

Things are somewhat more complicated if governments provide pure public goods and exclusion is not possible. The only possible equilibrium in this case is that none of the local governments provide the good. Since again, central provision will be inefficient with a Leviathan government, the welfare comparison must weigh two distortions, and general conclusions cannot be drawn.

The Leviathan model was originally proposed as a tool for developing guidelines for a fiscal constitution and less as a positive model of government decision making. As such, maximizing tax revenue is no doubt just as

extreme a government objective as welfare maximization. Most likely, government objectives are somewhere in the middle. Thus, government does not act without constraints, and in particular, needs to be reelected. Voter welfare will thus enter the objective function of government officials. Edwards and Keen (1996) model a government whose objective function is a combination of welfare and tax revenue. They show that in general, tax coordination is beneficial if the marginal excess burden of taxes (as percentage of the total marginal costs of taxation) is larger than Leviathan's marginal propensity to waste tax revenue.[23] This is rather intuitive: While a large marginal excess burden increases the benefit from reduced tax competition, this benefit is partially offset by government's waste of revenue. It is these two forces that citizens must weigh against one another when judging the desirability of tax competition versus tax coordination.[24]

But Leviathan is obviously only one example of political failure. Others may lead to spending below efficient levels. For instance, if old people without children play a decisive role in the political process, then spending on education may be too low from a social point of view. In this case, the downward pressure on tax rates caused by tax competition would be counterproductive.

In any case, tax competition is only a second best instrument to remedy political inefficiencies. The first best response would be a reform of political institutions. If this were not feasible for some reason, tax competition might be viewed as beneficial if those political inefficiencies are prevalent. We return to the problem of designing a constitution for local governments in Section 3.

2.2.4. Imperfect Mobility and Limited Number of Communities

Two important assumptions of the Tiebout model are that individuals are perfectly mobile and that there are enough communities offering distinct packages of public services such that each type of consumer finds a community which exactly satisfies her preferences. Clearly, both assumptions are too strong to be literally satisfied in reality. In the real world, moving is of course not costless. In addition to the physical act of moving which is costly, individuals may also be attached to their community if they have friends and family who live there; environmental amenities also may create a barrier to mobility. Therefore, individuals may find themselves stuck in a community with a non-optimal supply of public services. If government officials pursue their own private goals, such as maximizing tax revenue, an efficient allocation will not be reached with imperfect mobility (Brennan and Buchanan, 1980). In this case, competition cannot in itself limit governments' monopoly power. Citizens need to "regulate" government, as we will discuss in further detail in Section 3.

But for some purposes, the assumption of perfect mobility may be a close enough approximation. Thus, it has been suggested that the Tiebout model is best seen as a model for large metropolitan areas, where consumers can choose

at low costs among a number of different jurisdictions. But then, another problem becomes manifest, namely, there will typically not be enough jurisdictions for all types of consumers to acquire their preferred bundle of public goods. This means that some consumers of different types must live together. Since there will—in the absence of Lindahl taxes—generally be disagreement over policies, the mechanism of local public choice becomes important.

Public choice in local public goods models has been studied by a number of authors who analyze majority voting with mobile citizens (Ellickson, 1971; Westhoff, 1977; Epple, Filimon and Romer, 1984). In this section we present a simplified version of Westhoff's (1977) model to illustrate how these models work.[25] Suppose the economy consists of a given number of jurisdictions, which for simplicity we assume to be two. Individuals are perfectly mobile. They have identical quasiconcave utility functions $U(X, G)$, over a composite private good, X, and a public good, G. They differ, however, in their lump-sum income, y, which is distributed on some interval $[0, Y]$ according to a distribution function $F(y)$ with continuous density $f(y)$. Each jurisdiction levies a head tax, T, to finance provision of publicly provided goods. We will assume for now that the public good is a publicly provided private good, and that marginal costs are constant (normalized to one), so the cost function can be written $C(n, G) = nG$, where n is the population size consuming the good. Fiscal variables are chosen by majority rule.

It turns out that a majority voting equilibrium will exist if preferences satisfy the single crossing condition; in essence, the condition requires that voters' preferences over taxes and public goods can be ordered by some variable (here, income) independently of the policy adopted. In our model, a sufficient condition for this to hold is that public goods be normal. Mobility of individuals, coupled with the single crossing condition, implies that in equilibrium communities will be stratified by income: all individuals with income above (below) a certain threshold, call it $\tilde{y}$, live together in one community.

Westhoff (1977) proved existence of an equilibrium with a tax on exogenous income and pure public goods. He also found that stable equilibria are likely to exist only if there are multiple equilibria; if the equilibrium is unique, it is always unstable (Westhoff, 1979). Existence was also proven in a model with a housing market with property taxes and quasi-private goods by Epple, Filimon and Romer (1984, 1993).

Let us now turn to the question how voting affects the normative properties of an equilibrium, assuming it exists. The model above was used by Goodspeed (1998) to address this issue. He found that in general, local public choice will matter even in the case of head taxes and quasi-private goods. In the case where the equilibrium public good level is not optimal in both communities, the migration equilibrium will not be efficient either.

To see why this is the case, consider a two community model; individuals with income $y \in [0, \tilde{y}]$ live in community 1 and those with $y \in [\tilde{y}, Y]$ in community 2. To find an efficient allocation, maximize the sum of utilities of individuals living in the two communities:

$$\max_{T_1, T_2, \tilde{y}} \int_0^{\tilde{y}} U\left[y - T_1, T_1\right] f(y)dy + \int_{\tilde{y}}^{Y} U\left[y - T_2, T_2\right] f(y)dy.$$

The first order conditions can be written as

$$\int_0^{\tilde{y}} \frac{U_G^1}{U_X^1} f(y)dy = F(\tilde{y}), \tag{6}$$

$$\int_{\tilde{y}}^{Y} \frac{U_G^2}{U_X^2} f(y)dy = 1 - F(\tilde{y}), \tag{7}$$

$$U\left(T_1, \tilde{y} - T_1\right) = U\left(T_2, \tilde{y} - T_2\right). \tag{8}$$

While the last equation is satisfied by virtue of free mobility (this is the "boundary indifference condition"), the first two correspond to the Samuelson conditions for efficient public good supply: As is well known, within a jurisdiction, voting on public goods will produce a Pareto efficient allocation only when the marginal rates of substitution are distributed symmetrically about the mean (Bowen, 1943). This will hold only in particular cases, for instance, if the MRS is linear in income and the distribution is symmetric.[26] When equations (6) and (7) are satisfied, equation (8) shows that locational efficiency can be attained in an equilibrium. If, however, voting within communities leads to suboptimal choices, migration equilibria will generally be inefficient.

Before leaving this section, we should briefly recapitulate our discussion of the Tiebout model. As the literature has shown, the assumptions needed to guarantee an efficient equilibrium are very restrictive. Thus, distortionary taxation limits the efficiency properties of competition among governments. This opens up two different paths of dealing with the problem: One is the standard suggestion of tax coordination. The other would be to search for tax systems that more closely approach benefit taxation. Imperfect mobility and limited number of communities prevent the perfect sorting equilibria implied by the Tiebout mechanism. It is at this point that various imperfections become apparent: Governments may not be benevolent and citizens need to subject government to democratic or other controls. Even voting, though, suffers from well known problems related to the aggregation of preferences. In Section 3, we will come back to these problems from a constitutional angle: How can citizens ensure efficiency in the local public sector, if the Tiebout mechanism does not function perfectly.

2.3.   Empirical Tests of Local Public Goods Models

Tiebout showed that under restrictive assumptions, mobility of consumers leads to an efficient allocation of local public goods and an efficient allocation of consumers among communities. Therefore, whether or not the assumptions of the model hold is of considerable importance for policy recommendations. In this section, we briefly review empirical studies of local public goods models, focussing on two issues. The first set of studies is concerned with the publicness of locally provided goods. The second directly addresses the question of efficiency in public good supply.

As Bewley (1981) and others have shown, local provision of goods and services is most likely to be efficient when the goods are private in nature.[27] Empirically, the bulk of the studies have assumed that total provision, G, of a publicly provided good is related to the amount consumed by an individual, g, by the relation $G = N^\alpha g$, where α is a congestion parameter: for pure public goods, each individual consumes the entire amount provided, which implies $\alpha = 0$. Conversely, for pure private goods, each individual consumes one N-th of the total amount, i.e., $\alpha = 1$. This approach was used in two influential papers by Borcherding and Deacon (1972) and Bergstrom and Goodman (1973). Both studies assumed that the amount of public services provided corresponds to that demanded by the median voter. The parameter α can then be recovered from an estimation relating total expenditures to characteristics of the median voter. Both papers found that publicly provided goods were essentially like private goods.[28]

These findings have generated a literature too large to be reviewed here. We mention only a few of the criticisms against those early findings.[29] One important objection came from Oates (1988). He argued that larger towns typically provide a number of public goods and services for which a certain population threshold must be reached, such as zoos, theaters, museums, and so on. Hence, the finding that larger towns have higher expenditures may simply mean that they provide a greater variety of services, not necessarily that these services are congested. Apart from this, the measurement of the price of public services, the median voter's tax share, the functional form of the congestion function, are among the problems faced by researchers who estimate congestion functions. Moreover, whether the median voter model is really a good description of local politics is open to question.[30] Another issue is that the measurement of congestion should account for turnout (Borck, 2002): if some individuals abstain from voting, the decisive voter will not in general be the median income voter. For instance, suppose that turnout rises with income and falls with population size (see Borck, 2002 for a review of evidence). Then larger populations will have lower turnout, which means the decisive voter will be richer. If public good demand rises with income, this implies higher spending; however, the effect of increasing population on congestion comes partly from the

effect on turnout and not from congestion. All in all, while many of the studies have found publicly provided goods to be quasi-private, the last word on this issue is not spoken yet.

The discussion of the efficiency of local public good supply was started by Oates (1969). He observed that property values are positively related to the level of local public spending and negatively to the level of local taxation, thus confirming the Tiebout hypothesis of residents voting with their feet. The underlying idea is that housing prices reflect the discounted rents plus the value of public services less taxes. This approach has initiated a lively discussion on tax price, discount rate, functional form, the measurement of public services etc. (see Feld, 1999 for a summary). In particular, it has been argued that capitalization is inconsistent with the alleged efficiency properties of the Tiebout model. In equilibrium, public expenditures and revenues should typically not be related to housing prices. The value of a marginal unit of a public service should be reflected exactly in the marginal tax price. No capitalization should be observed. Capitalization will however occur if the Tiebout adjustment does not take place and the immobile capital bears a part or the full difference between public service and marginal willingness to pay.

Empirically, two groups of studies can be distinguished. One follows the simple question raised by Oates and merely asks whether capitalization takes place. Yinger, Börsch-Supan, Bloom and Ladd (1988), Stull and Stull (1991) and other scholars investigating U.S. local communities, and Kirchgässner and Pommerehne (1996) in their research on Switzerland, mostly found only partial capitalization of public expenditures as well as taxes. Feld (1999) shows for Switzerland that capitalization of taxes is much stronger for high than for low income recipients. A second group of studies has followed the approach of Brueckner (1982) to test directly for efficiency in public service provision. Brueckner showed that property values are an inverted U-shaped function of public service provision. He also showed that maximizing property values with respect to public service levels will lead to an allocation which satisfies the Samuelson condition for a Pareto optimal allocation. Thus, he suggested a simple test: If one finds that public goods do not influence property values, the null hypothesis of efficient public good supply cannot be rejected (under the assumption of property value maximization). Conversely, a positive (negative) coefficient in the regression of property values on public expenditure points to under-(over-) provision of public goods.[31] Brueckner (1982) found that in his sample of Massachusetts communities, public goods were neither systematically underprovided nor overprovided. Feld (1999) found that in Switzerland, public service provision is less than optimal in cities and close to the optimum in cantons (states). This would seem to be consistent with the intuition that tax competition leads to underprovision of public services, and this tendency is stronger at lower levels of government.

There have also been more direct tests of efficiency. Barlow (1970) used estimated demand functions to test whether provision of public education satisfies the Samuelson condition. He found that public education was generally underprovided. Bergstrom, Roberts, Rubinfeld and Shapiro (1988) repeated the test with survey data. They found that in their sample of Michigan school districts, public education seems to be provided efficiently.

Another test of interaction between jurisdictions with distortionary taxation has been developed recently: If each jurisdiction were "small," it should not react to policy changes by its neighbors. Strategic interaction, however, would occur, if jurisdictions are large enough to take neighbors' policies into account. While the verdict on this matter is still out, existing studies seem to indicate such interaction for specific policy areas (e.g., Case, Rosen and Hines, 1993; Brueckner, 1998; Buettner, 2001). For a survey of empirical studies, see Braeckner (2003),

3.   THE CONSTITUTIONAL APPROACH TO LOCAL GOVERNMENT

The discussion of the assumptions of the Tiebout model in Section 2.1 has shown that there are mainly two opposing arguments why competition among local governments may lead to inefficient results: destructive competition and monopoly power. The former is driven by economies of scale in consumption and from distortionary taxation—sometimes from their interaction—and the latter derives from limited choice of resident communities and from migration costs. Both inefficiencies require regulation. But the types of regulation to be applied are quite different. Destructive competition requires regulation to constrain excessive competitive forces, while regulation of monopoly power should neutralize forces intending to exploit citizens as taxpayers. Each of these regulations will be treated in turn. The idea of this section is to discuss alternative regulatory frameworks for local communities which can be considered as local government constitutions. The emphasis is on allocation, assuming that higher level governments take care of redistribution.

3.1.   Regulation of Destructive Competition among Local Governments

Here we assume that destructive competition operates and that the problem of monopoly power is absent. Destructive competition—sometimes dubbed "harmful competition"—is the main argument to harmonize local taxes and expenditures. The "regulatory authority" is the state or federal government or even a supranational government.

The question of the significance of destructive or harmful competition among local governments has been discussed at length in the literature (see

Sections 2.2.2 and 2.2.3 above). But it has hardly been investigated in the literature how regulation, if regarded as desirable, should be organized and how it would work.

Recall that the case for decentralized government made in Section 1.2 has been based mainly on two arguments: lower costs to adjust supply of public services to local preferences and lower costs of experimentation with innovation in the public sector. Regulation to prevent destructive competition will, however, require sacrificing these advantages to some extent. Regulation will require that the same principles be applied to all jurisdictions, while the logic of decentralization demands that public services be apportioned according to differing local preferences. Both goals are difficult to achieve simultaneously. This shows that regulation is available only at a cost which may exceed the benefits. Moreover, regulation may generate an endless chain of government interventions. For competitors will look for and find loopholes in the regulatory rules and therefore make new regulations necessary, and so on. In particular the expenditure side of local governments' budgets and local zoning regulations are open to circumventing competition.[32]

These problems may arise even if all participating governments are benevolent, i.e., if all local governments maximize the welfare of their citizens. When government officials pursue their own interests, however, such regulation may become even more problematic. For instance, revenue maximizing governments may seek for regulation in order to reduce competition beyond what is required to prevent destructive practices. Their goal may not be orderly competition, but a tax cartel guaranteed by central government's tax legislation. The central government may happily support such an endeavor if it can have its share of the increased revenue.

These short examples show that central government regulation intending to contain destructive competition among local jurisdictions may be counterproductive and may overshoot into cartel protection. Designing institutions which protect citizens and allow for efficient regulation is then of prime importance.

Alternatively, one may ask whether, in the absence of regulation, failures from destructive competition are likely to persist. Economics suggests that if there are unexploited gains of trade, competitive forces will in general find their way to realize them. If local taxes are distortionary as claimed by Zodrow and Mieszkowski (1986), one would expect that competition will find new forms of financing such as prices, access charges, and two part tariffs and benefit taxes to overcome the problems of destructive competition (Blankart, 2001; Blankart, 2002). The problem, from this perspective, is not too much competition, but the restriction of tax instruments available to local governments. A constitution should then not restrict local tax autonomy, except for the explicit prohibition of taxation which leads to violation of fiscal equivalence, e.g., in the form of tax exporting.

If public services exhibit economies of scale, competition may lead to inefficient results. But then, governments may delegate authority to some higher level. At some point, however, this will leave only one supplier, namely, the central government. The problem is then how to regulate a government monopoly.

3.2.   Regulation of Local Government Monopoly Power

Monopoly power of local governments (even in the absence of economies of scale and distortionary taxes) requires another type of regulation. If governments possess monopoly power due to immobility of residents or limited choice of communities, profit maximization by local governments as assumed in the original Tiebout model is no longer enough to bring about an efficient allocation. Citizens would be exploited by excessive taxation or insufficient public service levels. The traditional answer in existing state and federal constitutions is that local governments should be subjected to regulation by higher level governments. But it has been shown in the preceding section that such a regulation may be counterproductive if governments are selfish and maximize their power and revenue. An alternative is regulation by local residents. They have interests on the demand and supply side of public services. On the one hand, they are consumers of local public services, and on the other, they are citizens and as such owners of the facilities. Collectively, they can therefore be expected to have more encompassing interests than a politician acting as an agent of the central government. The task of regulating local governments could therefore best be conveyed to the local voters. They have an interest that charges and taxes for financing public services are neither monopolistically high nor insufficient to cover costs.

The idea of democratic self-regulation proposed here may be regarded as reasonable by most participants on the constitutional level. But the question is how local democracy should be organized. Should it be direct or representative democracy? In the first case collective decisions are taken by the voters in an open vote or in referenda, in the latter case, voters participate only via election of representatives. During the last 30 years an intense scholarly discussion has taken place on whether pure representative democracies or democracies with referenda and popular initiatives perform better. Most empirical studies present evidence from the United States and Switzerland, since only in these two countries is direct democracy frequently observed. Particular attention has been given to local governments, and it is there where institutional variety is greatest. But the discussion encompasses also the state and federal levels (for a survey see Kirchgässner, Feld and Savioz, 1999).

From a normative point of view, it would be desirable to test directly for the efficiency of direct versus indirect democracies. This is, however, a rather

ambitious goal, and, therefore, hardly any researchers have directly addressed this question. In a study of referenda in the canton of Basel-Stadt between 1950 and 1977, Noam (1980) analyzed whether the referenda led to inefficient results. He found that in all but 5 % the referendum outcome passed the compensation test, i.e., the winners could have compensated the losers.[33] The only direct test of allocative efficiency that we are aware of is by Deller and Chicoine (1993). They perform a test of efficiency in public good supply using Brueckner's approach (see Section 2.3), and find no systematic difference between direct and indirect democracies. More evidence along these lines is clearly needed.

Most of the literature has asked the less ambitious question whether institutions matter. There are basically two types of studies in this category. One strand considers the effect of democratic institutions on various performance measures. The other type, to be treated below, tries to ascertain whether direct democracy produces results that are closer to the median voter's preferred allocation than representative democracy.

Several studies have addressed government spending in direct versus indirect democracies. For Switzerland, Kirchgässner, Feld and Savioz (1999) show that expenditures are lower in cities with referendum on budgetary matters than in those without referendum. Matsusaka (1995) studies the effect of the initiative on expenditures in the U.S. states. He finds that, while combined state and local government spending is lower in states with the initiative, local government expenditures are higher in states with the initiative.

In these studies, the question whether in direct democracies, expenditures are closer to voters' preferences is not addressed. From the studies that do directly pose this question, the answer seems to be yes. Pommerehne (1978), using data from Switzerland, showed that the median voter model explains government expenditures much better in direct democracies than in representative ones. Gerber (1996, 1999) and Gerber and Hug (1999) concentrate on particular issues such as parental consent laws for abortion in case of minors, minority rights for homosexuals and the adoption of capital punishment. They show that the median preferences for these issues—which are derived from survey data—are more closely followed in states with the institution of an initiative than in states without such an institution.

The fact that democracies with direct voter participation come closer to median voter preferences would, however, not yet be enough to convince a critic of the desirability of direct democracy. For one thing, it has been clear since Bowen's (1943) seminal contribution that the median voter outcome is not in general Pareto efficient (see, also, Section 2.2.4). Moreover, one might argue that majoritarian decisions through direct democracy do not allow for Pareto superior decisions attainable through negotiations within the

small numbers of parliamentary representatives. Therefore, lower expenditures per capita in direct democracies might not be regarded as a sign of efficiency, but of inefficiency in the form of unexploited Pareto improvements. This argument is, however, not fully convincing. First, all issues subject to popular vote are in general discussed in the parliament before they come to a vote. In Switzerland, governments frequently propose their own alternatives as a compromise to a popular initiative. Voters then decide on three alternatives: the popular initiative, the government compromise, and the status quo. Hence, bargaining and popular vote are not necessarily mutually exclusive, but rather complementary.[34] Second, it is well known that bargaining as such is not always Pareto improving as the proponents of the representative democracy model suggest. Under majority rule, it often leads to logrolling outcomes which may be Pareto-inferior compared to no decision at all.

Finally, there are other indicators for direct democracy performing more efficiently than pure representative democracy. Feld and Savioz (1997) estimated cantonal production functions by adding a dummy variable for referenda on tax rates and deficits. Cross section analysis for 1989 as well as panel analyses for 1982-1993 show productivity levels which are higher by 15 and 5 percent respectively under direct than under representative democracy (other things equal). Thus, it seems that direct control over tax rates and deficits by voters induces governments to pursue a more growth oriented budgetary policy.

Another performance indicator is tax evasion. Weck-Hannemann and Pommerehne (1989) and Feld and Frey (2002) have shown that tax evasion decreases as more tax payers are able to participate in the political decision making process through referenda, because transparency and perceived fairness of a tax system increase with more direct democracy. Hence, the propensity of the tax payers to follow the rules of the game increases. Weck-Hannemann and Pommerehne (1989) show that the right of the voters to vote on the level of public service output, on tax rates, and on net deficits negatively affects the amount of income tax evasion. Feld and Frey (2002) show that in more directly democratic cantons tax authorities show more trust in front of taxpayers. In direct democracies, the authorities do not automatically suspect taxpayers to be cheating. Moreover, sanctions for minor violations are lighter and sanctions for major violations larger than in representative democracies, which may imply more trust but also use of sanctions to enforce an implicit contract between the authorities and the taxpayer.

4.   CONCLUSION

In this essay, we have reviewed some of the literature on local public finance. At the heart of this field has traditionally been the question whether local policies lead to an efficient allocation of resources. The Tiebout model

claims they do, which has aroused severe criticism both on theoretical and empirical grounds. We have argued that one lesson from failure of the Tiebout mechanism is that local public choice needs to be modelled explicitly. Barriers to mobility and limited choices of communities lead to failure of perfect stratification as envisaged by Tiebout. Policy outcomes then depend on the institutions of public choice, and these institutions clearly matter.

In closing, we would like to emphasize some areas of research we think are not yet well developed. While the literature is full of Tiebout type models, analysis of political institutions in local policies is rather scant. In comparing different political institutions, it appears that there is a need for a clearer link between theory and empirics. On the one hand, some of the empirical literature which studies "performance" lacks a clear theoretical underpinning. Just like finding of the flypaper effect led to development of political models, findings of institutional differences may lead to the development of more formal models of local politics. While direct democracy is relatively easy to apply to local public finance, representative democracy is not, since agency problems abound, and these should be linked to other features of the local system, e.g., mobility of people and factors of production. On the other hand, some of the formal literature is still awaiting the formulation of precise theoretical predictions which lend themselves to empirical testing. Therefore, local public finance seems to be a field where scholars may still reap large payoffs. We hope that this will cause further studies which advance our understanding of local politics.

NOTES

1. See also Borck (2002) who argues that participation is not only a goal in itself but is important for its allocative effects as well. Since in general, rich people tend to participate more in politics, increasing participation by making jurisdictions small should raise political participation by the poor. The effects on redistribution and allocation should be taken into account.
2. Breton (1965) calls this "perfect mapping," and Oates (1972), "perfect correspondence." A different criterion is the principle of institutional congruency proposed by Blankart (2000). It requires that the collective decision making process is organized in such a way that the sets of those who benefit, of those who pay the taxes and of those who make the collective decision coincide. The intention is to avoid external costs of decisions made by outside governments.
3. The applicability of the Coase Theorem presumes that (i) there are no transactions costs and bargaining agreements can be costlessly enforced, (ii) preferences are common knowledge, (iii) bargaining agents perfectly represent their constituencies, and (iv) the parties will agree to a division of the surplus (Inman and Rubinfeld, 1997b, p. 76).
4. Oates (1999) reviews some efforts at coordinated action by state or local governments in the U.S. It appears that in those cases where cooperation has been successful (e.g., the management of the Chesapeake Bay and the creation of an Ozone Transport region), coordination was facilitated or initiated by the federal government.

5. For an overview of intergovernmental grants, see Gramlich (1977), and more recently Oates (1999).

6. There are also some arguments that unconditional grants may be needed to correct locational inefficiencies (Boadway and Flatters, 1982).

7. Empirical studies of intergovernmental grants have found, however, that grants are not well explained by normative predictions but rather by political factors, such as the representation of local jurisdictions in higher level parliament or the political affiliation between local and higher level politicians (e.g., Inman, 1988 and Borck and Owings, 2003).

8. See, e.g., Fisher (1982), Oates (1994), and Hines and Thaler (1995) for surveys.

9. Similar conclusions follow if interest groups succeed in diverting funds away from the median voter's preferred allocation (Dougan and Kenyon, 1988).

10. More formally, one needs to assume that there are at least as many distinct communities as there are types of individuals. Additionally, there must be enough consumers of each type such that the optimal community size can be attained.

11. See Comes and Sandler (1996) and Scotchmer (2002) for surveys of club models.

12. Even with pure public goods, however, the optimal community size may be finite, since crowding occurs due to decreasing returns to labor or increasing commuting costs implied by larger communities.

13. The result that mixing is inefficient stems from the absence of complementarities in the production of public goods. In the presence of such complementarities, complete stratification is no longer efficient (Schwab and Oates, 1991).

14. In Bewley's example, this is an equilibrium because he assumes "democratic" governments who do the best they can for citizens living within their jurisdiction. This would clearly not be an equilibrium with "entrepreneurial" governments who would want to attract citizens.

15. Scotchmer also assumes that maximizing property values can be achieved through property taxes on land value. If the property tax is a distortionary tax on capital, the results change (see Section 2.2.2).

16. The problem was already discussed by Oates (1972). Modern theoretical discussion started with Wilson (1986) and Zodrow and Mieszkowski (1986). For good surveys, see Wildasin (1986), Wellisch (2000), and Wilson (1999).

17. Since citizens are identical and a symmetric equilibrium will be assumed, we drop the jurisdiction specific subscripts.

18. See Wilson (1999) for a survey.

19. Wellisch (1995) shows that a tax on mobile capital leads to an efficient supply of local public factors if the tax can be viewed as a user fee. This is the case if the use of the public factor is proportional to capital input.

20. If regions are small—so that they take the return to capital as given—then the median voter will never rely on capital taxes in equilibrium (Borck, 2003; Fuest and Huber, 2001). This matches results on the use of tax mix in a framework where governments maximize the welfare of their citizens (Bucovetsky and Wilson, 1991).

21. In fact, Hamilton showed that zoning transforms the property tax into a head tax, for the following reason: A household cannot consume less housing than the zoning ordinance dictates, and if it consumed more, it could move to another community with a higher minimum level of housing. Since this community could provide the same public good level with a lower tax rate, the move would make this household better off. Thus, in equilibrium, each household within any jurisdiction consumes the same amount of housing, and the property tax is not distortionary.

22. This ignores a number of issues. For instance, if the central government is a Stackelberg leader vis-a-vis the states, it may be able to set its tax rate so as to achieve a first best allocation.

23. The excess burden comes from outmigration of the mobile factor in response to a rise in tax rates.

24. For a similar approach using voting, see Fuest and Huber (2001). They obtain similar results for the desirability of tax coordination, but the political distortion induced by tax coordination now stems from overprovision of public goods, if median income is below average income.

25. Epple, Filimon and Romer (1984) extend Westhoff's model to include a housing market.

26. This last statement, in fact, requires that the income distribution be symmetric in *both jurisdictions, in equilibrium.* Obviously, a symmetric national distribution is not sufficient for this.

27. The finding that publicly provided goods are private goods would seem to suggest that the supply of these goods could be privatized. But there may be various sorts of inefficiencies under private supply, such as asymmetric information, externalities, and so on. Thus, the finding that the public sector provides private goods does not necessarily imply these goods should be provided by the private sector.

28. While Borcherding and Deacon (1972) used state-level data, Bergstrom and Goodman (1973) used data from New York communities. The articles also differed somewhat in the assumptions, e.g., on the tax price faced by voters.

29. The interested reader is referred to Reiter and Weichenrieder (1997) for a useful survey.

30. Pommerehne (1978) studied Swiss communities and distinguished between those with direct and representative democracy. While the results clearly matter for the performance of the median voter model, no clear pattern emerges for the measurement of congestion.

31. Brueckner notes, however, that the efficiency results are only second best, since with a property tax, housing choices will be distorted, and the Samuelson rule then implies efficiency conditional on the existing (inefficient) housing stock.

32. In this vein, Cremer and Gahvari (2000) show that in the presence of tax evasion, tax coordination alone is not sufficient; rather, countries also need to coordinate their audit policies.

33. However, while this is evidence that the projects approved were not inefficient—i.e., projects with negative net social benefits passed or projects with positive net social benefits rejected—it does not prove that the projects were also of efficient size.

34. In a recent contribution, Matsusaka and McCarty (2001) present a model of the initiative where there are agency problems. Politicians propose a policy, which becomes law unless an interest group makes a counter proposal (the initiative). In some circumstances, the threat of the initiative may cause politicians to propose extreme policies which makes voters worse off. But initiatives are valuable in that they increase the set of choices available to voters, and if agency problems are not severe, the initiative always benefits the voter.

REFERENCES

Barlow, R. (1970). "Efficiency Aspects of Local School Finance." *Journal of Political Economy* 78: 1028-1040.

Bergstrom, T. C. and Goodman, R. (1973). "Private Demands for Public Goods." *American Economic Review* 63: 280-296.

Bergstrom, T. C., Roberts, J. A., Rubinfeld, D. L., and Shapiro, P. (1988). "A Test for Efficiency in the Supply of Public Education." *Journal of Public Economics* 35: 289-307.

Bewley, T. F. (1981). "A Critique of Tiebout's Theory of Local Expenditures." *Econometrica* 49: 713-740.

Blankart, C. B. (2000). "The Process of Government Centralization." *Constitutional Political Economy* 11: 27-39.

Blankart, C. B. (2001). "Taxes and Choice. Two Views on the Regulation of Tax Competition." *Jahrbuch für Neue Politische Ökonomie* 19: 129-143.

Blankart, C. B. (2002). "A Public Choice View of Tax Competition." *Public Finance Review* 30: 366-376.

Boadway, R. and Flatters, F. R. (1982). "Efficiency and Equalization Payments in a Federal System of Government: A Synthesis and Extension of Recent Results." *Canadian Journal of Economics* 15: 613-633.

Borcherding, T. E. and Deacon, R. T. (1972). "The Demand for the Services of Non-federal Governments." *American Economic Review* 62: 891-901.

Borck, R. (2002). "Jurisdiction Size, Political Participation, and the Allocation of Resources." *Public Choice* 113: 251-263.

Borck, R. (2003). "Tax Competition and the Choice of Tax Structure in a Majority Voting Model." *Journal of Urban Economics* 54: 173-180.

Borck, R. and Owings, S. (2003). "The Political Economy of Intergovernmental Grants." *Regional Science and Urban Economics* 33: 139-156.

Bowen, H. R. (1943). "The Interpretation of Voting in the Allocation of Resources." *Quarterly Journal of Economics* 58: 27-48.

Bradford, D. F. and Oates, W. E. (1971). "The Analysis of Revenue Sharing in a New Approach to Collective Fiscal Decisions." *Quarterly Journal of Economics* 85: 416-439.

Brecht, A. (1932). "Internationaler Vergleich der Öffentlichen Ausgaben." *Grundfragen internationaler Politik,* 2.

Brennan, G. and Buchanan, J. M. (1980). *The Power to Tax.* Cambridge: Cambridge University Press.

Brueckner, J. K. (1982). "A Test for Allocative Efficiency in the Local Public Sector." *Journal of Public Economics* 19: 311-331.

Brueckner, J. K. (1998). "Testing for Strategic Interaction Among Local Governments: The Case of Growth Controls." *Journal of Urban Economics* 44: 438-467.

Brueckner, J. K. (2003). "Strategic Interaction Among Governments: An Overview of Empirical Studies." *International Regional Science Review* 26: 175-188.

Buchanan, J. M. (1965). "An Economic Theory of Clubs." *Economica* 32: 1-14.

Bucovetsky, S. and Wilson, J. D. (1991). "Tax Competition with Two Tax Instruments." *Regional Science and Urban Economics* 21: 333-350.

Buettner, T. (2001). "Local Business Taxation and Competition for Capital: The Choice of Tax Rate." *Regional Science and Urban Economics* 31: 215-245.

Case, A. C., Rosen, H. S., and Hines, J. R., Jr. (1993). "Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States." *Journal of Public Economics* 52: 280-296.

Cornes, R. and Sandler, T. (1996). *The Theory of Externalities, Public Goods, and Club Goods,* 2nd ed. Cambridge: Cambridge University Press.

Cremer, H. and Gahvari, F. (2000). "Tax Evasion, Fiscal Competition and Economic Integration." *European Economic Review* 44: 1633-1657.

Deller, S. C. and Chicoine, D. L. (1993). "Representative versus Direct Democracy: A Test for Allocative Efficiency in Local Government Expenditures." *Public Finance Review* 21: 100-114.

Dirlmeier, U. (1966). "Mittelalterliche Hoheitsträger im wirtschaftlichen Wettbewerb." *Vierteljahresschrift für Sozial- und Wirtschaftsgeschichte,* (Beiheft Nr. 51).

Dougan, W. R. and Kenyon, D. A. (1988), "Pressure Groups and Public Expenditures: The Flypaper Effect Reconsidered." *Economic Inquiry* 26: 159-170.

Edwards, J. and Keen, M. (1996). "Tax Competition and Leviathan." *European Economic Review* 40: 113-134.

Ellickson, B. (1971). "Jurisdicitonal fragmentation and residential choice." *American Economic Review* 61: 334-339.

Ennen, E. (1972). *Die europäische Stadt des Mittelalters,* Göttingen: Vandenhoek & Ruprecht.

Epple, D., Filimon, R., and Romer, T. (1984). "Equilibrium Among Local Jurisdictions: Toward an Integrated Treatment of Voting and Residential Choice." *Journal of Public Economics* 24: 281-308.

Epple, D., Filimon, R., and Romer, T. (1993). "Existence of Voting and Housing Equilibrium in a System of Communities with Property Taxes." *Regional Science and Urban Economics* 23: 585-610.

Feld, L. P. (1999). Steuerwettbewerb und seine Auswirkungen auf Allokation und Distribution: Eine Empirische Analyse für die Schweiz. Ph.D. thesis, Universität St. Gallen.

Feld, L. P. and Frey, B. S. (2002). "Trust Breeds Trust: How Taxpayers Are Treated." *Economics of Governance* 3: 87-99.

Feld, L. P. and Savioz, M. R. (1997). "Direct Democracy Matters for Economic Performance: An Empirical Investigation." *Kyklos* 50: 507-538.

Filimon, R., Romer, T., and Rosenthal, H. (1982). "Asymmetric Information and Agenda Control: The Bases of Monopoly Power in Public Spending." *Journal of Public Economics* 17: 51-70.

Fischel, W. A. (1992). "Property Taxation and the Tiebout Model: Evidence for the Benefit View from Zoning and Voting." *Journal of Economic Literature* 30: 171-177.

Fisher, R. C. (1979). "A Theoretical View of Revenue Sharing Grants." *National Tax Journal* XXXII: 173-184.

Fisher, R. C. (1982). "Income and Grant Effects on Local Expenditure: The Flypaper Effect and Other Difficulties." *Journal of Urban Economics* 12: 324-345.

Fisher, R. C. (1996). *State and Local Public Finance,* 2nd ed. Chicago: Irwin.

Frey, B. S. and Eichenberger, R. (1999). *The New Democratic Federalism For Europe. Functional, Overlapping and Competing Jurisdictions.* Cheltenham, U.K.: Edward Elgar.

Fuest, C. and Huber, B. (2001). "Tax Competition and Tax Coordination in a Median Voter Model." *Public Choice* 107: 97-113.

Gerber, E. R. (1996). "Legislative Response to the Threat of Initiatives." *American Journal of Political Science* 41: 99-128.

Gerber, E. R. (1999). *The Populist Paradox; Interest Group Influence and the Promise of Direct Legislation.* Princeton: Princeton University Press.

Gerber, E. R. and Hug, S. (1999). *Minority Rights and Direct Legislation: Theory and Methods.* Mimeo: University of California at San Diego.

Gilbert, G. and Picard, P. (1996). "Incentives and Optimal Size of Local Jurisdictions." *European Economic Review* 40: 19-41.

Goodspeed, T. J. (1998). "On the Importance of Public Choice in Migration Models." *Economics Letters* 59: 373-79.

Gramlich, E. M. (1977). "Intergovernmental Grants: A Review of the Empirical Literature." In *The Political Economy of Fiscal Federalism,* Oates, W, E., ed., pp. 219-239. Lexington, Mass.: D.C. Heath.

Hamilton, B. W. (1975). "Zoning and Property Taxation in a System of Local Governments." *Urban Studies* 12: 205-211.

Hamilton, J. H. (1986), "The Flypaper Effect and the Deadweight Loss from Taxation." *Journal of Urban Economics* 19: x148-55.

Hartmann, P. C. (1978). "Die Steuersysteme in England und Frankreich am Vorabend der Französischen Revolution." In *Vom Ancien Régime Zur Französischen Revolution,* Hinrich, E., Schmidt, E., and Vierhaus, R., eds., pp. 43-65. Göttingen: Vandenhoek & Ruprecht.

Hines, J. R. and Thaler, R. H. (1995). "Anomalies: The Fflypaper effect." *Journal of Economic Perspectives* 9: 217-226.

Hoxby, C. M. (1996). "Are Efficiency and Equity in School Finance Substitutes or Complements?" *Journal of Economic Perspectives* 10: 51-72.

Inman, R. P. (1988). "Federal Assistance and Local Services in the United States: The Evolution of a New Federalist Fiscal Order." In *Fiscal Federalism: Quantitative Studies,* Rosen, H. S., ed., Chicago: Chicago University Press.

Inman, R. P. and Rubinfeld, D. L. (1997a). "Making Sense of the Antitrust State-Action Doctrine: Balancing Political Participation and Economic Efficiency in Regulatory Federalism." *Texas Law Review* 75: 1203-1299.

Inman, R. P. and Rubinfeld, D. L. (1997b). "The Political Economy of Federalism." In *Perspectives on Public Choice,* Mueller, D. C., ed., pp. 73-105. Cambridge: Cambridge University Press.

Keen, M. (1998). "Vertical Tax Externalities in the Theory of Fiscal Federalism." *International Monetary Fund Staff Papers* 45: 454-485.

Kirchgässner, G., Feld, L. P., and Savioz, M. R. (1999). *Die Direkte Demokratie.* Basel: Helbing & Lichtenhain.

Kirchgässner, G. and Pommerehne, W. W. (1996). "Tax Harmonization and Tax Competition in the European Union: Lessons from Switzerland." *Journal of Public Economics* 60: 351-371.

Matsusaka, J. G. (1995). "Fiscal Effects of the Voter Initiative: Evidence from the Last 30 Years." *Journal of Political Economy* 103: 587-623.

Matsusaka, J. G. and McCarty, N. M. (2001). "Political Resource Allocation: Benefits and Costs of Voter Initiatives." *Journal of Law, Economics, and Organization* 17: 413-448.

McGuire, M. (1974). "Group Segregation and Optimal Jurisdictions." *Journal of Political Economy* 82: 112-132.

McLure, C. E. (1967). "The Interstate Exporting of State and Local Taxes: Estimates for 1962." *National Tax Journal* 20: 49-77.

McLure, C. E. (1986). "Tax Competition: Is What's Good for the Private Goose Also Good for the Public Gander?" *National Tax Journal* 34: 341-348.

Mieszkowski, P. (1987). "Urban Economics." In *The New Palgrave Dictionary of Economics,* Eatwell, J., Milgate, M. and Newman, P., eds., pp. 756-762. London: McMillan.

Mieszkowski, P. and Zodrow, G. R. (1989). "Taxation and the Tiebout Model: The Differential Effects of Head Taxes, Taxes on Land Rents, and Property Taxes." *Journal of Economic Literature* 27: 1098-1146.

Musgrave, R. A. (1959). *The Theory of Public Finance.* New York: McGraw-Hill.

Neenan, W. B. (1970). "Suburban-Central City Exploitation Thesis: One City's Tale." *National Tax Journal* 23: 117-139.

Noam, E. M. (1980). "The Efficiency of Direct Democracy." *Journal of Political Economy* 88: 803-810.

Oates, W. E. (1968). "The Theory of Public Finance in a Federal System." *Canadian Journal of Economics* 1: 37-54.

Oates, W. E. (1969). "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis." *Journal of Political Economy* 77: 957-971.

Oates, W. E. (1972). *Fiscal Federalism.* New York: Harcourt Brace Jovanovich.

Oates, W. E. (1979). "Lump-sum Intergovernmental Grants Have Price Effects." In *Fiscal Federalism and Grants-in-Aid,* Mieszkowski, P. and Oakland, W. H., eds., pp. 23-30. Washington, D.C.: Urban Institute.

Oates, W. E. (1988). "On the Measurement of Congestion in the Provision of Local Public Goods." *Journal of Urban Economics* 24: 85-94.

Oates, W. E. (1994), "Federalism and Government Finance." In *Modern Public Finance,* Quigley, J. M. and Smolensky, E., eds., pp. 126-151. Cambridge, Mass.: Harvard University Press.

Oates, W. E. (1999). "An Essay on Fiscal Federalism." *Journal of Economic Literature* 37: 1120-1149.

Oates, W. E. and Schwab, R. M. (1988). "Economic Competition Among Jurisdictions: Efficiency Enhancing or Distortion Inducing?" *Journal of Public Economics* 35: 333-354.

Olson, M. (1969). "The Principle of 'Fiscal Equivalence': The Division of Responsibilities Among Different Levels of Government." *American Economic Review* 59: 479-487.

Pack, J. R. (1991). "The Opportunities and Constraints of Privatization." In *Privatization and its Alternatives,* William, J. and Gormley, T., ed., pp. 281-306. Madison: University of Wisconsin Press.

Pommerehne, W. W. (1978). "Institutional Approaches to Public Expenditure: Empirical Evidence from Swiss Municipalities." *Journal of Public Economics* 9: 255-280.

Pommerehne, W. W. and Krebs, S. (1991). "Fiscal Interactions of Central City and Suburbs: The Case of Zurich." *Urban Studies* 28: 783-801.

Popitz, J. (1927). "Der Finanzausgleich." In *Handbuch der Finanzwissenschaft,* pp. 338-375. Tübingen: Mohr.

Reiter, M. and Weichenrieder, A. (1997). "Are Public Goods Public? A Critical Survey of the Demand Estimates for Local Public Services." *Finanzarchiv* 54: 374-408.

Rubinfeld, D. L. (1987). "The Economics of the Local Public Sector." In *Handbook of Public Economics,* Auerbach, A. J. and Feldstein, M. S., eds., Vol. 2, pp. 571-645. Amsterdam: North-Holland.

Samuelson, P. A. (1954). "The Pure Theory of Public Expenditure." *Review of Economics and Statistics* 36: 387-389.

Schwab, R. M. and Oates, W. E. (1991). "Community Composition and the Provision of Local Public Goods." *Journal of Public Economics* 44: 217-237.

Scotchmer, S. (1994). "Public Goods and the Invisible Hand." In *Modern Public Finance,* Quigley, J. M. and Smolensky, E., eds., pp. 93-119. Cambridge, Mass.: Harvard University Press.

Scotchmer, S. (2002). "Local Public Goods and Clubs." In *Handbook of Public Economics,* Auerbach, A. J. and Feldstein, M., eds., Vol. 4, pp. 1997-2042. Amsterdam: Elsevier.

Stull, W. J. and Stull, J. C. (1991). "Capitalization of Local Income Taxes." *Journal of Urban Economics* 29: 182-190.

Tautscher, A. (1947). *Staatswirtschaftslehre des Kameralismus.* Bern: Francke.

Tiebout, C. M. (1956). "A Pure Theory of Local Expenditures." *Journal of Political Economy* LXIV: 416-424.

Tolley, G. S. (1974). "The Welfare Economics of City Bigness," *Journal of Urban Economics* 1: 324-345.

Tullock, G. (1969). "Federalism: Problems of Scale." *Public Choice* 6: 19-29.

Weck-Hannemann, H. and Pommerehne, W. W. (1989). "Einkommensteuerhinterziehung in der Schweiz: Eine Empirische Analyse." *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* 125: 515-56.

Wellisch, D. (1995). *Dezentrale Finanzpolitik bei hoher Mobilität.* Tübingen: Mohr.

Wellisch, D. (2000). *The Theory of Public Finance in a Federal State.* Cambridge: Cambridge University Press.

Westhoff, F. (1977). "Existence of Equilibrium in Economies with a Local Public Good." *Journal of Economic Theory* 14: 84-112.

Westhoff, F. (1979). "Policy Inferences from Community Choice Models: A Caution." *Journal of Urban Economics* 6: 535-549.

Wildasin, D. E. (1986). *Urban Public Finance*. Amsterdam: Harwood Academic Publishers.

Wildasin, D. E. (1989). "Interjurisdictional Capital Mobility: Fiscal Externality and a Corrective Subsidy." *Journal of Urban Economics* 25: 193-212.

Wilson, J. D. (1986). "A Theory of Interregional Tax Competition." *Journal of Urban Economics* 19: 296-315.

Wilson, J. D. (1999). "Theories of Tax Competition." *National Tax Journal* 52: 269-304.

Yinger, J., Börsch-Supan, A., Bloom, H. S., and Ladd, H. F. (1988). *Property Taxes and House Values: The Theory and Estimation of Intrajurisdictional Property Tax Capitalization.* San Diego: Harcourt Brace Jovanovich.

Zodrow, G. R. and Mieszkowski, P. (1986). "Pigou, Tiebout, Property Taxation and the Underprovision of Local Public Goods." *Journal of Urban Economics* 19: 356-370.

Chapter 17

FEDERALISM AND SUBSIDIARITY IN NATIONAL AND INTERNATIONAL CONTEXTS

Jean-Michel Josselin
Université de Rennes 1 and Centre National de la Recherche Scientifique
jean-michel.josselin@univ-rennes1.fr

Alain Marciano
Université de Reims Champagne Ardenne and Centre National de la Recherche Scientifique (GREQAM) and IDEP
a.marciano@wanadoo.fr

Abstract	A mainly European and continental notion, subsidiarity is nevertheless at the heart of federalism. Either in an international context or in order to federate a nation, the economics of constitutions provides a nuanced view of the merits of the different forms of federalism. It builds on the tools of economic analysis to try to draw some lessons from the American experience that could serve to better understand the ongoing institutional changes in Europe.
Keywords:	Agency theory, American constitution, Cameralism, competitive institutions, constitutional dilemma, European Union, federalism, Subsidiarity
JEL classification:	H10

1.   INTRODUCTION: GREAT EXPECTATIONS

Are constitutions an object of economic investigation? In 1887, Woodrow Wilson proposed to close down "the business of manufacturing constitutions" (quoted by Ostrom, 1976, p. 1). The only task left would be the management of administration. The constitution-making process would be completed, the constitution manufactured and there would be no more interest in deepening investigations. Thus, the institutional box could be closed and therefore become a black box. In this perspective, the next question would then consist

in examining individual behaviors within a given constitutional framework. It is the approach adopted by public choice scholars. Of course, to recognize, as is emphasized by Frey (1990) and North (1990), that institutions do indeed matter and that no individual could boast total emancipation from them, is necessary. Nevertheless, one should go one step beyond. Indeed, we observe the working of constitutions in practice and at the same time face constitutional moments, since there is an ongoing process of mergers and dissolution of nations (Wittman, 1991; Blum and Dudley, 1991; Alesina and Spolaore, 1997; Bolton and Roland, 1997). The reason is also to be found in the very nature of constitutions and especially in their function. Aranson insists that a constitution does not only serve a managerial but also a value function. "It does so in a particular way, involving uncertainty about the time-bound goals of the process to be managed and about what values are permissible outcomes from the process so managed" (Aranson, 1988, p. 293). Since the constitutional design is never completed, it is necessary to go inside the black box and to complement the public choice analysis with a constitutional political economy perspective (Buchanan, 1990). In the latter case, the question of choice within rules is replaced by the question of the choice of the rules themselves. Constitutional political economy does not only state that economic analysis cannot but take constitutions seriously, it also affirms that constitutions are susceptible to economic analysis.

The constitutional problem first of all concerns the choice between the unitary structure and the federal organization. The following developments will focus on the political economy of federalism, which will not prevent comparisons with unitary systems. The first approach to federalism is political, mainly with Montesquieu, dating back to an era when economics did not exist as a separate and autonomous discipline. We shall try to show that economic tools help decipher the rationale for and the various forms of federalism. The latter is indeed everything but monolithic, from confederacies to the different organizations of federal powers. The objective here is to make out these different meanings.

The constitution-making process in Eastern Europe, the evolution of the Australian institutions, the European integration and enlargement decisions, devolution and the status of regions and states, all inform us of the real-world problems that must be analyzed. They also evidence the need for a political economy approach that complements the standard study of fiscal federalism. While the latter has been given widespread and thorough attention for quite a long time (Oates, 1972, 1998, 1999; Inman and Rubinfeld, 1996), the political economy of federalism is a relatively more recent subject of interest (Inman and Rubinfeld, 1997a, b), although Buchanan and Tullock's *Calculus of consent* does make a seminal case for an economic theory of constitutional

choice. Though we shall use some aspects of fiscal federalism, our focus is on the economics of the constitution of the federal state itself.

Whether unitary of federal, constitutions are charters of government. As such, they define how the collective decision-making process is organized, how and how far power has to be divided among different entities. The major difference between those two forms of organization is that a unitary constitution delegates power to a single authority, while a federal one intends to share public prerogatives among several institutions, none of them granted with the ultimate power. Thus, the decisive question raised by federalism is how to assign responsibilities, powers, rights or prerogatives to each of the different levels of government. To put it differently, even if the word has rarely been used at least until its introduction in the semantics of the European Union, subsidiarity is the key concept to enter the world of federalism. It allows to identify a federal constitution according to the following criteria. Firstly, are responsibilities primarily assigned to the upper or to the lower level? Secondly, how do jurisdictions compete or bargain when deciding the assignment of prerogatives to a given level? Thirdly, how is the mechanism of control of the federal construction designed? Subsidiarity basically refers to the primary delegation of tasks to the smallest functional unit (Backhaus, 1999). Then, "in the clear case of insufficiency of a particular level, the nearest functional one needs to be found" (Backhaus, 1997a, p. 138). Following the intuitions of Christian Wolff (see Dreschler, 1999, for a biography), if a given level of government or province cannot adequately solve a problem, it can first turn to another province, in which case subsidiarity is lateral. If cooperation is searched with a higher level, then upwards subsidiarity prevails. Finally, downwards subsidiarity expresses that the problem is handled by communities of a lower level inside the province.

For all these theoretical considerations, federalism develops in a historical background that is by no means neutral. To federate a nation, as was the case in the eighteenth century in North America, does differ from federating already existing nations, as it occurs in the European integration process. Our demonstration mainly builds on the constitutional history of the U.S.A. and on the current situation in the European Union. Both cases can be considered as archetypal illustrations of federal systems, of their virtues and their limits.

The chapter is organized as follows. In the Section 2, we present a model of the federal state based on the principal-agent theory, intending to clarify the link between subsidiarity and federalism. Then, we analyze them respectively in national contexts (Section 3) and in international contexts (Section 4).

2.   THE ECONOMIC APPROACH TO FEDERALISM

Like so many institutions, a federation, with its many facets, is a creation of human activity. As such, its legitimacy can and must be questioned. The first

step is to ask whether we need federalism at all in order to efficiently organize societies (2.1). Once this relative efficiency is assessed, the second step consists in describing and evaluating the different forms of federalism (2.2). Even if at this stage, the discussion remains largely theoretical, we must bear in mind that these are actual and very acute debates in practice. For instance, in West Africa, Cameroon was a federal state in the sixties (Constitution of the 14th of August 1961) but decided afterwards to become a unitary state (Constitution of the 20th of May 1972).

2.1.   Why Federalism?

Once the conditions of the necessity of the state are spelt out, there remains the question of the form that people want to give to it. Should it be a unitary state granting the power to a single authority? Should power be fragmented among several authorities? We will have to describe the economic framework from which answers to these questions can be put forward. An economic model of government is proposed (2.1.1) which is then used to characterize a unitary organization (2.1.2) and a federal system (2.1.3).

2.1.1.   State of Nature or Constituted State?   Do we really want a constituted state? Cannot we do without it? After all, spontaneous order can be a very effective way of organizing societies, as is exemplified by the not so wild "wild west" experience (Anderson and Hill, 1977). However, conditions of Pareto-efficiency in a spontaneously ordered society are quite strong and individuals may want to construct a constituted state in order to get out of the state of nature (2.1.1.1). This move to an artificially ordered society cannot avoid the question of the nature of the corresponding constitutional contract. Agency theory may be a fruitful analytical basis from which to analyze constitutions (2.1.1.2) and make out their specificity (2.1.1.3).

2.1.1.1.   Getting out of the State of Nature   If the state of nature is Hobbesian, then the outcome of human interactions is doomed to be Pareto-inefficient. We obviously must get out of this prisoners' dilemma. If on the contrary, one can contend that the state of nature is Humean, then we need not get out of it, since it is naturally pacific. Sympathy may help players to choose a Pareto-efficient outcome. However, Hume himself warned us against any over-enthusiasm: Societies governed by sympathy are usually very small (Josselin and Marciano, 1999b, 2001) and are not likely to be replicated at larger scales. If the scope of human interactions is ever to expand, we must then consider ways of getting out of a probably inefficient state of nature.

The usual means of creating artificially ordered societies are either conventional or contractual. In the first setting, tradition and the progressive learning

of rules shape conventions. In the second, there is an explicit discussion of the goals that will be reached collectively. Federalism falls into this category: "The key issue was whether societies of men could create good government from reflection and choice" (Ostrom, 1976, p. 3). Conventional or contractual, the state is constituted, in the first case by a set of conventions, in the second case by a constitutional contract.

2.1.1.2.   The Nature of the State: Constitutions as Agency Contracts

Constitutions can be viewed as agency contracts by which people delegate to constituted public institutions the right and duty to undertake some given tasks. There exists a large body of literature that uses the principal-agent framework to describe legal or constitutional issues (see for instance, Kau and Rubin, 1979; Peltzman, 1980, 1984; Kalt and Zupan, 1984; Anderson and Hill, 1986; Wagner, 1988, 1993; Aghion and Tirole, 1997; Qian and Weingast, 1997; Cooter, 2000; Josselin and Marciano, 2000). Hobbes himself depicted the allocation of power and authority as an activity of delegation, of representation, thus using the very terms of the agency theory. In this setting, constitutions are agency contracts. Citizens, the people, or any emanation from them, delegate to a ruler the task of organizing a Pareto-improving social order, as a consequence of the necessity to get out of the state of nature. Here the principal is the people and the ruler is the agent. If citizens choose to grant delegation to a single agent, then the state is unitary. If several agents are involved, then the state is federal. To write down a constitution or to provide constitutional conventions requires comprehending all these features in a contract that is necessarily specific.

2.1.1.3.   Specificity of the Constitutional Agency Contract   The first

specificity of constitutions as agency contracts is the nature of the constitutional incompleteness. Usually, most incomplete contracts are such that projects cannot be fully described: "[t]he nature of projects' payoffs is initially unknown to both the principal and the agent" (Aghion and Tirole, 1997, p. 6). Nevertheless, the number of projects is generally assumed to be finite and known (Merville and Osborne, 1990; Aghion and Tirole, 1997). However, delegation of power becomes even more elaborated when the set of projects is not necessarily finite.

Because of this incompleteness, the contract cannot give the fullest possible description of the set of projects. Control of the agent(s) is then a critical issue, all the more that shares of government are not necessarily alienable. The second specificity of constitutional contracts indeed comes from the nature of shares of ownership in government. Contrary to agency contracts in profit-seeking organizations, there is no capital market for governments. Control is therefore more difficult since selling or buying shares to optimize a portfolio is

everything but straightforward. "[...] capital markets generally do not operate directly in the case of governments [...] Governments, as they are presently constituted, are non profit cooperatives in which ownership shares are inalienable. People cannot buy and sell shares of ownership in government, so these ownership rights cannot acquire a capital value [...] The absence of a market for ownership shares in governments makes it impossible for people to specialize in ownership. People cannot acquire multiple shares of ownership; they acquire ownership by virtue of residency and must relinquish that ownership upon changing residency" (Wagner, 1988, pp. 428-429).

The first way of dealing with the specificity of the constitutional contract is to design a unitary structure.

2.1.2.   Unitary Constitutional Contracts   The first way of defining a constitutional contract consists in delegating power to a single agent. This is the perspective of unitary states: Property rights on government shares are not alienable (2.1.2.1). At the same time, it enhances the direct and clear-cut responsibility of the single agent (2.1.2.2), but at the price of a monopolist power that generates a constitutional dilemma (2.1.2.3).

2.1.2.1.   A Single Hobbesian Agent: Inalienable Shares of Government   In the case of a unitary social contract, citizens are as many principals granting all the power to their agent. Unitary states can then be defined by a unique ultimate authority. In this perspective, citizens cannot acquire multiple shares of ownership. Furthermore, their own property rights cannot be sold. In other words, shares in government ownership cannot be alienated. Indeed, having delegated power to the ruler, they cannot but obey him since any defection would be a breach to the contract and a threat of return to a Hobbesian state of nature.

2.1.2.2.   A Constitutional Monopolist and Central Assignment of Prerogatives   The mechanism of delegation requires a strictly and formally defined constitutional control of the state. In the unitary case, the contract intends to reduce incompleteness by positively defining—authorizing or prohibiting—the prerogatives granted to the single agent. It thus delineates a set of possible projects. This formal definition, and delimitation, of actions helps provide the agent with the final authority. Consequently, the agent is sovereign: He is outside the law he makes for the players of the post-constitutional game. There is thus a monopoly power over the making of rules.

The sovereign agent (these are quite contrasted terms!) of course delegates in his turn tasks or activities, but retains the formal authority. If the sovereign power were to submit to law then it would no longer be sovereign. Any further delegation of tasks is a matter of efficiency (scale effects for instance), not of

division of power: This is simply decentralization and it fundamentally differs from federalism. As Bish (1988, p. 360) puts it: "Analyses of federalism that ignore the monopoly problem and focus only on scale problems are indistinguishable from analyses of decentralized unitary government system." Delegation of tasks is different indeed from delegation of power: The former does not affect the sharing of ownership while the latter builds on it. Decentralization is thus not a relevant key to distinguish unitary from federal states. What matters is that, in unitary states, the central government controls the assignments and the possible re-assignments (Salmon, 1987). This is a constitutional prerogative: If powers are unilaterally assigned by the highest level, then the state is unitary, even if it can be very decentralized. The constitutional assignment of powers to lower tiers of government may be quite significant in some unitary states, for instance in France, but the central government controls the procedure for the revision of the constitution. There is no competitive assignment, no possible unilateral action by lower levels of government to change the allocation of powers.

A surrender of power to a subordinate government that is complete, permanent, and of "constitutional magnitude" comes under the framework of devolution (Tannenwald, 1998). Any institution that intends to put forward this kind of reform must of course possess a sovereignty of its own, or at least a delegated sovereignty. For instance, the British parliament does have this capacity. It has been using it quite recently in Scotland and it progressively extends the movement to other regions. More generally, unitary states have by construction an effective power to devolve responsibilities to lower levels of government. Prerogatives granted to the latter are of a constitutional nature. It may be a way of breaking the monopoly power that expresses the constitutional dilemma of unitary contracts.

2.1.2.3. **Control of the Agent and the Constitutional dilemma** The constitutional dilemma arises because of an infinite regress in the control of the controllers (who guards the guardian?). The contract may prevent it since it defines an *ex ante* control of the state. Any attempt at *ex post* control would fall into inefficient outcomes as in Samaritan's dilemma or inspection games. The primary concern lies with the control of the monopoly power to make law that rests with the state (Josselin and Marciano, 1997). If the provision of security follows increasing returns then such a natural monopoly is justified, as is acknowledged by Nozick (1974). Regulation may help the state combat discretionary behaviors of its agents but the sovereign power itself cannot be contained by introducing competitive mechanisms. Indeed, only internal control is conceivable: the ruler "delegates" to himself the right to be judged (Josselin and Marciano, 1995; Cooter, 2000). This is the essence of administrative law in contractual unitary states, as is the case in France. Let us note that in coun-

tries ruled by constitutional conventions, like England, the impartial spectator of the common law judges the actions of the state. The framework is still unitary but the sovereign power of the ruler is not established by a contract but rather by conventional norms (Josselin and Marciano, 1999a).

The actions of the sovereign come under contractual positive assignments, with norms of judgement using the French tradition, very influential for instance in continental Europe. Judgements are rendered according to substantive and *"quant au fonds"* (as to the content) rules. Control is thus exercised by rules rather than by discretion. It provides gains in predictability but leads to losses in flexibility (Cooter, 2000, Chapter 6). Organizing "a government through constitutional design is preeminently a planning process that purposefully may suppress procedural spontaneity" (Aranson, 1988, p. 296). Aranson provides this comment with respect to the federal American constitution but it obviously applies to unitary constitutions as well. What remains a significant point is the value of flexibility when societies face ongoing movements and changes.

2.1.3.   Federal States: Competition between Agents   The second approach to the constitutional contract is that of a federal state. It consists in delegating power to several agents. Therefore, public prerogatives are assigned through competition, horizontal as well as vertical, between agents sharing the power granted to them (2.1.3.1). People can still relinquish ownership, but under strict conditions and at high costs (2.1.3.2). The reverse of the medal is the possible collusion between agents and the dilution of liabilities (2.1.3.3)

2.1.3.1.   A Competitive Assignment of Responsibilities   What first characterizes federal states is the absence of any agent gathering under his name the ultimate authority. Power is delegated by "the people," to use the words of Madison, in such a way as to fragment public prerogatives and to divide sovereign power among many agents. This is the system of checks and balances described by Madison and Hamilton in the Federalist Papers. No single agent has the final word. States, the federal government and other components such as parliament, are put into competition, presumably to the best advantage of the principal (even though their authority is necessarily always overlapping, they participate in the checks and balances of the post constitutional game). Whatever its form, the ruling power is to remain the agent of the people, which is the case, for instance, in the present American constitution, as we will see later.

The constitution here does not positively define prerogatives but rather organizes the conditions of competition between the different agents. However, it is not only horizontal competition between agents that matters. Their vertical distinction is also of great importance. The definition of prerogatives is

endogenously provided by this competition. This is certainly in sharp contrast with the unitary perspective where incentives to refrain from Leviathan-like behaviors are as tightly delineated as possible. "Most discussions of constitutional *limits* seek to impose some enforced passivity on politicians, as in seeking to restrain them to some substantive listing of enumerated powers or in seeking to preclude them from engaging in some lists of activities. But for a society of risen apes, a more acceptable constitutional framework would seem to be one that not so much seeks to impose passivity, as it allows fully active politicians, limited only by the natural forces of competition, *rightly constructed"* (emphasis added) (Wagner, 1993, pp. 54-55). Active competition between agents is of course enhanced by the ability of citizens to alienate their shares of government.

2.1.3.2.   Competition between Agents and Alienable Ownership Shares of Government   Ownership of government is acquired by virtue of residency (Wagner, 1988). Agents of lower levels, states or regions for instance, come under this competitive mechanism by which individual mobility, even if it is not without cost, can prompt governments into efficiency. Devolution too can enhance competition by giving lower tiers of government new constitutional prerogatives. The federal constitution must of course allow it beforehand. At the national level, ultimate government ownership has the same essence as in unitary states. It nonetheless remains that agents of the same as well as of different levels will compete for the approval of citizens. Whilst remaining in the federation, individuals can at a cost move from one state to another, thereby contributing to the increase of diversity and emulation among states.

2.1.3.3.   Control of the Federation and the Constitutional Dilemma In the unitary perspective, control of the state is ex ante insofar as it takes place at the writing of the constitution. In the federalist approach, control is ex post and ongoing (it keeps being redefined) since it rests on competition between agents: "For the Federalists, the control of factions lay in the creation of an extended or compound republic, through which the position of final authority would be abolished, and replaced with a network of countervailing constitutional guardians" (Wagner, 1993, p. 50). Once the constitutional contract is agreed upon, then the sovereign power is separated between the agents but the constitution remains an incomplete parchment (the Hobbesian contract would in counterpoint attempt to establish a definite list of prerogatives, the problem being the internal means of precluding the single ruler from infringing them). Agents in the federal contract are granted with the responsibility to complete the parchment. Ongoing competition also requires that the document remain unsealed so that the rulers will be in capacity to deal with future developments

and events. This competition may be stable, to the benefit of the principal, the people. It can also degenerate either into collusion or into monopoly. In the first case, "Although federalism is designed to provide "substitute" governments so that the officials of different government or branches of a government watch each other, there is nothing intrinsic in federalism that would prevent collusion among government officials or between government officials and some citizens" (Bish, 1988, p. 357). Secondly, competition may turn to the advantage of one of the agents who monopolizes power. In both cases, the power of the principal is at best attenuated.

The risk is at least as important as it is in a unitary state. An agent in dominant position may try to reverse the agency relationship and behave as if he were the principal. Since in the federalist system there is no ex ante list of prohibited or authorized actions, since the now monopolist agent decides what they are, then nobody guards this guardian. In the case of collusion, the problem is of a similar nature but it may be less acute if cartels of power are as unstable as they are on industrial markets. Control of the state may thus be quite loose in the absence of formally defined authority that could be set as the ultimate locus of final liability. In practice, the federal government seems to acquire a somewhat pervasive influence, becoming the main "guardian." That could be explained in economic terms, since it is involved in vertical competition only whereas states are also engaged in horizontal rivalry. Judicial control is mainly the task of the court granted with the power to interpret the constitution, possibly at the expense of the states but to the benefit of the federal level. Any court that could acquire the authority to say what the constitution is would claim this role. We shall later have to assess this intuition when considering real-world constitutional systems.

Both unitary and federal governments require constitutional provisions in order to impose limitations upon the authority of officials (Ostrom, 1976). Control of the state takes two related forms. The first one considers possible economic predation: This must be impeded by constitutional provisions that guarantee and promote economic liberty for citizens. The second one should secure and enhance political participation in order to bridle the temptations of democratic Leviathans. In this respect, flexibility in the design of the federal contract ensures that new problems or cases will be dealt with in the most efficient manner. At the same time, flexibility will facilitate the ongoing re-negotiation of public prerogatives. Reciprocal confidence between individuals and politicians should result therefrom: The constitution becomes self-enforcing (Weingast, 1997). The second form of control rests on competition between agents, both horizontally and vertically. Federalism is nevertheless prone to collusion or monopolization: To what extent can it then promote economic as well as political liberty? The next step of our discussion consists in asking how to organize horizontal as well as vertical competition between

agents, in other words how to assign responsibilities and prerogatives among them.

2.2.   Which Federalism? Federalism and Subsidiarity

Federalism may take different forms but nonetheless rests on the general basis of the Tiebout model of local competitive institutions (2.2.1). However, Tiebout assignments are not without limits (2.2.2). Federalism must go beyond this first step by building further institutional settings, each based on a different acceptation of subsidiarity (2.2.3).

2.2.1.   A Common Ground for Federalism: Tiebout Assignment for Local Public Goods   The Cameralists (Backhaus and Wagner, 1987; see also Chapter 1) first developed this view of competitive institutions, later systematized by Tiebout (1956) and Bewley (1981). This model will not be formally described here (see Chapter 18). The focus will rather be on the consequences of this economic criterion on the organization of public prerogatives and the assignment of responsibilities between levels of government (2.2.1.1). However, Tiebout assignments both require strong assumptions and a careful consideration of the underlying market analogy (2.2.1.2).

2.2.1.1.   Locals Know and do Best (in Some Circumstances)   The principle is that congested and excludable public goods should be provided at the local level. The government of highest level (be it unitary or federal) should assign the responsibility of providing such public goods to local governments. Subsidiarity "optimize[s] the performance of the larger political entity, primary liability for the solution of problems lies with the smallest functional unit" (Backhaus, 1997a, p. 138). It provides an economically efficient assignment of authority. Firstly, this assignment ensures a "close match between the functions of a particular unit and the means and responsibilities to serve this function" (Backhaus, 1999, p. 136). Secondly, the argument according to which "locals know best" is reinforced by the competition between utility-taking jurisdictions or clubs. Using a game theoretic interpretation, the situation can be described as a non cooperative but at the same time non conflicting game. In other words, the relevant framework for describing this competition would not be a prisoners' dilemma.

Citizens or consumers have the ability to move from one jurisdiction to another according to their preferences and the cost of mobility. This *exit* option also influences the behavior of the suppliers of public goods. Breton and Scott (1978) suggest that competition should be preferred to cooperation since it prevents collusion and reduces transaction costs between jurisdictions. It should also prevent high levels of taxation (Weingast, 1995). Therefore, competition

is likely to reduce the discretionary use of power and opportunities of preda-
tion as well as profit opportunities for governments. Mobile factors have to
be attracted and thus should not be too heavily taxed while the burden on im-
mobile factors should be limited by the constraint of re-election. The "market
metaphor" (Donahue, 1997, p. 74) intends to explain how "[j]ust as market
competition pressures firm managers to reflect the interests of shareholders,
competition among local governments helps to limit government's predatory
behavior" (Qian and Weingast, 1997, p. 88).

Such competition also provides incentives to better control public firms by
avoiding "soft budget constraints" (Kornai, 1976). The bailout of inefficient
projects increases the opportunity cost of public expenditures. Qian and Wein-
gast (1997) argue that federalism induces governments to commit to negative
market incentives that punish public inefficiency while at the same time it lim-
its state predation and thus enhances entrepreneurship. The Cameralist setting
of competitive state enterprises was very much in the same vein. Neverthe-
less, its efficiency as well as its relevance was largely dictated by historical
instances, providing a sort of real-world laboratory.

2.2.1.2.   The Market Metaphor and Jurisdictional Competition   If
it were perfectly efficient, the principle of competing jurisdiction could alone
fully define federalism. However, any generalization of Tiebout assignments
may be risky since it builds on the "analogy from jurisdiction to market prod-
uct" (Musgrave, 1997, p. 67). Cameralist scholars seem to have been clearly
aware of solving a given problem in given historical and institutional circum-
stances. As is suggested by Donahue (1997, p. 75), "[t]he conditions hold well
enough, for a wide range of transactions, to amply justify popular and schol-
arly enthusiasm for market competition. But extending the logic to government
competition requires far more conceptual leaps."

Providing collective services on a territorial and competitive basis is a very
appealing way of building federalism, particularly when we bear in mind the
efficiency of market competition. Subsidiarity appears as a strong basis on
which to ground federalist principles. However, this basis is both imperfect and
incomplete. Tiebout competition requires too heavy assumptions to correctly
match the workings of competitive markets. Moreover, it may not be able to
cover the whole range of public services.

**2.2.2.   The Limits of Tiebout Assignments: The Grounds for Extend-
ing the Realm of Federalism**   The realm of federalism goes beyond the
scope of the Tiebout model for at least two reasons. Firstly, conditions for the
model to be effective are quite strong: "The market in which jurisdictions play
is decidedly imperfect" (Musgrave, 1997, p. 67). Jurisdictional competition it-
self, in some instances which define the intrinsic limits of Tiebout assignment

of public prerogatives, may impair economic efficiency (2.2.2.1). Secondly, even if jurisdictional competition works, it does not embrace all the aspects of collective action. The careful listing provided by Tiebout of the conditions of validity of his model could appear at first glance as punctilious. It is nevertheless the soundest basis on which to draw the limits of the approach. Some of the assumptions underlying the competitive model may indeed be too strong to fully depict any institutional reality. It then happens that many economic interactions between governments and individuals cannot be comprehended in the Tiebout model (2.2.2.2).

2.2.2.1.   Intrinsic Limits of Jurisdictional Competition   Obviously if competition corresponds to the non cooperative and not conflicting situation described in the preceding paragraph, positive effects should result from this process. However, if the institutional framework in which jurisdictional competition takes place is non cooperative and conflicting, it may have the characteristics of a prisoners' dilemma. This competition will necessarily lead to a result that coincides with what theory predicts, the worst possible outcome. Just as excessive market competition may generate price wars and may be detrimental to economic efficiency, jurisdictional competition has its drawbacks (Zodrow and Mieszkowski, 1986).

Competition may generate strategic behaviors from local governments. Production of law at the state level can induce strategic behaviors by voters: Legal externalities may help them catch the benefits of a program and transfer the costs onto the residents of other states. Examples may be taxes exported to non residents or investment attraction at the expense of other states by tax exemption or corporate regulation. In this matter, competition on corporate charters is absent in unitary systems. In federations, mobility of course enhances this competition and may offset the possibly negative effects of strategic behaviors. But mobility is costly and some players may be tempted by short term positive gains. "In the short run, a jurisdiction may benefit from that course, but others may feel forced to follow and a 'race to the bottom' may ensue" (Musgrave, 1997, p. 69). When considering the nation as a whole, the level of public services is not optimal. Competition generates this shortcoming.

2.2.2.2.   Extrinsic Limits   Even if we assume that Tiebout competition efficiently provides local public goods, it nonetheless remains that it cannot cover the whole range of public policies. The Tiebout setting cannot comprehend pure public goods or spillovers.

Tiebout's model explicitly deals with public goods that are excludable and subject to some rivalry. Formulated as an answer to Samuelson (1954), it complements the Lindahl-Bowen-Samuelson solution to the provision of pure public services. Tiebout's theory of jurisdiction size prefigures Buchanan's more

general theory of clubs (Buchanan, 1965). By definition, it cannot provide any institutional setting for Samuelsonian public goods. Exclusion mechanisms and preference revelation through mobility (Sandier and Tschirhart, 1997) cannot be simply replicated. Federalism nevertheless has to tackle this question and we shall consider later examples of how crucial it has been in the development of an actual federation.

The Tiebout model relies on a perfectly elastic supply of jurisdictions. No transaction costs hinder the capacity to construct purposeful clubs. So unrealistic an assumption certainly neglects the fact that the making of jurisdictions often has "more to do with the accidents of a capricious history than with the shifting dictates of economic rationality" (Donahue, 1997, p. 75). The U.S.A. offers an immediate but obviously not unique instance of a federation whose development "did not follow the blueprints of spatial efficiency" (Musgrave, 1997, p. 66). Therefore, the supply of jurisdictions is everything but perfectly elastic since their number, boundaries and to a lesser extent their population change very slowly (or, in economic terms, at great cost).

As a consequence, the market metaphor does not apply so straightforwardly. Local governments are no longer competitive suppliers but each government considers "itself as a Cournot competitor, aware of the potential effect of its action on other jurisdictions" (Inman and Rubinfeld, 1997b, p. 82). Furthermore, heterogeneity in the "elasticity of location" (Donahue, 1997, p. 77) of individuals or imperfect and differentiated mobility increase those effects and all the more reduce the scope of Tiebout competition. For instance, when workers leave one jurisdiction for less productive employment in another fiscally more attractive local community, the consequence is a less efficient private economy (Buchanan and Goetz, 1972; Boadway and Flatters, 1982). To put it differently, when the limits and the number of jurisdictions is fixed or at least very stable, the negative effects of competition are likely to be strengthened, thereby reinforcing the race to the bottom.

Limits of Tiebout assignments of responsibilities are twofold. They first relate to the possible defects of jurisdictional competition, just as market competition may degenerate if for instance the framework is that of a prisoners' dilemma. Second, the realm of collective action goes far beyond the provision of local public goods, which requires an extension of the realm of federalism as well.

2.2.3.   Extending the Realm of Federalism

When a given province, to use the words of Wolff, does not efficiently carry out its task, it can turn to three solutions. Firstly, downwards subsidiarity can restore economic efficiency. Either spatial or functional, lower levels of government inside the province may well be assigned rights that previously were prerogatives of the

province (2.2.3.1). Secondly, a federal constitution can be built on a bargaining game between neighboring provinces. Negotiations among them organize the assignment of the provision and financing of "national" public goods. Lateral subsidiarity provides the framework for confederation. The upper level of government remains the endogenous and not necessarily stable outcome of this ongoing bargaining (2.2.3.2). Thirdly, the federal government may also be granted a constitutionally defined set of prerogatives, namely finding ways of overcoming spillovers and providing national public goods. Upwards subsidiarity is the logical consequence of the expected efficiency of the federal level (2.2.3.3).

2.2.3.1.   Downwards Subsidiarity: Assigning Power to Lower Levels of Government

A first form of downwards subsidiarity is typically exemplified by devolution. This transfer of functions or prerogatives from a superior government to a lower tier (Kincaid, 1998) must be permanent and of constitutional rather than administrative nature. Otherwise, it would only be a matter of decentralization. The rationale for giving "regions" constitutional powers has many facets. (Drèze, 1993) explores some of them as he tries to assess the advantages and drawbacks of a regional status that would untie the link with the state while reinforcing it with the highest level of government. Drèze takes the possible example of Europe but the scope of his discussion is much wider. The idea is to evaluate the capacity of the reform to bring about Pareto improvements in terms of risk sharing, mutual insurance and its effects on incentives to increase regional productivity. Another way of considering the constitutional status of regions is to explicitly take space into account. As was masterly demonstrated by Hotelling (1929), geographical and preference spaces are intertwined and this simple fact generates strong interdependencies between locational and political decisions. Among others, Alesina and Spolaore (1997), Bolton and Roland (1997) or Josselin and Marciano (1999c) use a spatial framework to assess the relative merits and shortcomings of regional autonomy.

A second form of downwards subsidiarity is provided by the concept of functional federalism (Casella and Frey, 1992). The corresponding concept of Federal Overlapping and Competing Jurisdictions (Frey, 1996; Frey and Eichenberger, 1995, 1996, 1997a, 1997b) is an original answer to imperfect mobility and the relatively inelastic supply of geographical jurisdictions. This lack of elasticity is partly due to the weight of history, the slowness of institutional change, and the associated transaction costs. FOCJs view competition between institutions providing collective services on a functional rather than territorial basis. Jurisdictional competition à la Tiebout is replaced by functional competition. There is no longer a necessary link between ownership of

governmental shares (understood here as shares in the firm or the organiza-
tion that provides a given collective service) and residency. These functional
forms of government can of course keep a territorial basis, for instance they
can cover ecological entities such as water basins or mountain ridges. They
can also exist without a territorial link, such as when a profession is publicly
chartered and publicizes the corresponding rules and duties attached to it.

The principle underlying the concept of FOCJs is thus that of downwards
subsidiarity. Competition remains the incentive towards efficiency but it no
longer has to be organized on a geographical basis. The Tiebout model finds
here an important extension, even if we must bear in mind the limits of the
market metaphor.

**2.2.3.2.   Lateral Subsidiarity: Assignment of Powers as a Bargain-
ing Process**   The idea here is that the states will bargain in order to define
which public goods the federal representation must provide. In many respects,
a confederation is an ongoing process of negotiation about the prerogatives
that will be granted to itself. A shared goal of course enhances the benefits of
lateral subsidiarity. A straightforward example of it is the necessity to provide
a common defense for the fledgling United States of America. In economic
terms, to implement the shared objective amounts to providing a public good,
which is of course the case with defense. The first step of the process of con-
federation is thus a matter of bargaining about the level and financing of such
a good. Getting together and joining forces is one thing, providing the ma-
terial conditions to fulfil this goal is quite another one. Lateral subsidiarity
cannot dispense with considering the expenditure side and the revenue side
of the common project. The ensuing bargaining amounts to a discussion in
the Coasian spirit: Inman and Rubinfeld (1997b) give a detailed and insightful
interpretation of confederacies in these terms. This bargaining is unstable by
nature. Even if the states have agreed upon a given level of public services
that the federal representation must provide, the very nature of the agreement
makes its enforcement problematic. As long as the union keeps a confederate
form, the federal level will not be granted with compelling prerogatives. But
then, the sharing of the cost of the public good comes under usual problems of
free-riding. Authoritarian mechanisms may be effective but they would change
the nature of the initial agreement, transferring so much power to the federal
representation that the agency relationship may be substantially modified, to
the detriment of the principals. Incentive-compatible mechanisms that could
prevent free-riding would better suit the spirit of the confederation, but the
translation of quite sophisticated economic methods in constitutional terms
may be tricky. One may also hope that as the public good provision game is
repeated, even the most hardened egoists will learn sympathy, which brings
us back to Hume. However, establishing "a firm league of friendship" (to use

the wording of the Articles of Confederation, 1781, Article III) may not be so straightforward a task, as Hume himself warned.

All this makes the confederation process an ongoing bargaining which may not be without advantages. For instance, fluctuating objectives due to changing circumstances may require this flexibility. Furthermore, since lateral subsidiarity rests on similar objectives and constraints for the joining states, preference revelation may not be too acute a problem. However, free-riding prevention and enforcement of cost-sharing decisions cannot be settled so easily. The plasticity of the confederate contract may ensure potential responsiveness but not necessarily actual decisiveness. Hence the temptation of a more authoritarian federalism.

2.2.3.3. Upwards Subsidiarity: Assigning Power to Higher Levels of Government

The assignment of prerogatives to the federal government is first of all justified on the basis of economic efficiency (Oates, 1972). While a confederation relies on an ongoing process of bargaining about the prerogatives that are granted to the agents, a federation rests on a constitutionally given delimitation of the competencies of every level of government. Upwards subsidiarity is justified by the existence of pure public goods or spillover effects. This is so because it is assumed that the federal level is efficient whilst the local level is not. However, the assumption does not necessarily hold since the efficiency of the national government is questionable. The main reason is that when the assignment of prerogatives is fixed, there is no competition in the provision of public services at the federal level, hence the possible bureaucratic or rent-seeking behaviors that have been emphasized by public choice scholars.

On the other hand, the assignment of prerogatives between levels of government is not as straightforward as it may seem. The preceding assumption must therefore be lifted, or at least carefully assessed. Two examples may illustrate this point. Firstly, education is usually considered as a local matter. However, externalities in the formation of human capital due to the mobility of workers are an incentive to provide education at the national level. Secondly, is income distribution a national or a local problem? For instance, Pauly (1973) considers that a local treatment seems to better match individual preferences. However, differences in taxation among jurisdictions may induce changes in the allocation of the population. Mobile individuals with high revenues may try to avoid taxes by voting with their feet whilst people with low income will be attracted by important levels of public services. The magnitude of these effects depends on the elasticity of location which was previously referred to. Some morally questionable practices can remedy the consequences of those strategic behaviors without going to the root of the initial inefficient provision

of collective services. For instance, if the latter is to remain a local preroga-
tive, then the strategic mobility of poor people must be precluded. That was
the case in England from the Act of Settlement of 1662 until 1795. Similarly,
The American Articles of Confederation precluded this kind of mobility. Pro-
vision of services at the national level would not of course require so drastic
a measure and would possibly restore economic efficiency. Furthermore, any
attempt to reduce mobility lessens the scope of the Tiebout setting.

Another form of upwards subsidiarity is that of metropolitan governments
(Hochman, Pines and Thisse, 1995). This is a way of extending and renewing
the framework of federalism by putting emphasis on the spatial constraints in
organizing large urban areas. In some respects, Tiebout's model may not be ge-
ographical enough. Transport costs are absent, both between communities and
inside them. Hochman, Pines and Thisse (1995) then propose a model that de-
picts federalism on a territorial rather than functional basis. Metropolitan gov-
ernments provide the whole range of public services to the individuals residing
in their jurisdictions. The usual institutional setting in which public goods are
provided by as many functional clubs is not efficient when space is explic-
itly taken into account. Profit-maximizing clubs will get a revenue from user
charges that will not cover the cost of provision of the service. A metropolitan
government will ensure that provision costs for the whole range of goods is
covered by user charges and the aggregate land rent on the territory. There is
thus no problem of rent-sharing or of spillovers among spatially overlapping
functional clubs. Resting on the Henry George rule (see Starrett, 1988), local
governments with a large enough territorial basis will maximize land-value.
Ownership shares of government are explicitly linked to residence. This up-
wards shift puts an emphasis on the revenue side of subsidiarity. Inefficient
rent-sharing mechanisms are avoided and the metropolitan government has a
clear incentive to an efficient provision of services since it increases land rent.
However, there may be by construction only a small number of metropolitan
areas. Local government may then no longer be utility takers, which departs
from the Tiebout idea.

2.3.   Conclusion of Section 2

What shows through all the previous developments is the question of sta-
bility. This is a difficult one since stability does not have a single acceptation.
It can be a matter of secession or break-up of nations. In the agency frame-
work used here, it stems from the paradox of government: Delegation of power
necessarily brings about several forms of contest. A first one may be the mo-
nopolization of the delegated power by one of the agents, even if the agency
relationship itself remains stable. In the purely competitive case (in Tiebout or
FOCJ settings), no such monopoly threats exist. A second form of instability

may come from the possibility of delegation reversal, where the principal is displaced by its agents or by the most influential of them. In the latter case, monopolization of delegated power and threat of reversal are closely related. A third form of disturbance arises whenever the assignment of the respective and reciprocal rights and duties of the states and of the federation is a subject of contention. In the terms of the agency theory, it amounts to conflicting claims as to who the principal is. The history of the American federalism provides striking instances of all these questions, which may provide as many lessons for forthcoming federations, and possibly for Europe.

3. FEDERATING A NATION: SUBSIDIARITY IN NATIONAL CONTEXT

When entering the realm of federalism, one of the main problems is the definition of rights and prerogatives of the agents. At the same time, this brings about interrogations as regards the principalship. Is the people the principal? Or do the states naturally play this role above individuals? All this could be purely theoretical speculation, were it not for the obvious practical dimension of the distinction. The U.S.A. give a sometime dramatic illustration of it. If a federal government goes against the interests of a given state, can it be the case that individuals belonging to this state will feel bound to their primary roots and contest the provisions of the national government? This is of course the premonition of Tocqueville. In economic terms, some individuals may want to abandon their principalship to their state. The federal position then becomes tricky. The Hamiltonian vision is that the national government is the agent of the people. The interest of the latter should transcend that of particular states. In other words, states should not be considered as principals since this situation would convey the risk of factions that could claim principalship in contradiction to the permanent and general interest of the community. This is of course the view of Madison and the Federalists. We give here an economic interpretation of what amounts to be a competition for principalship. In the first period of the confederacy, the states remain the principals but lateral subsidiarity very quickly shows its limits (3.1). Federalism is then intended to bring principalship back to the individuals but it nonetheless raises many questions as far as the prerogatives of the agents are concerned (3.2).

3.1. Providing Public Goods through a Confederacy: The American Example

The American confederation is a remarkable example of quite novel ideas put into practice (3.1.1). The attempt establishes principles of lateral subsidiarity (3.1.2) that will however be somewhat shaken by strategic behaviors (3.1.3).

3.1.1.   The Economics of "a Firm League of Friendship"   Adopted
in 1781, the Articles of Confederation are the first constitution of the United
States. Previous attempts to design constitutional frameworks in the American
colonies had failed. For instance, the Albany Plan of Union of 1754, approved
by the Albany Congress, was never ratified by the other colonial legislatures.
All the states already had their own constitution and their own government.
What was at stake was to gather them under a unified national setting. Since
the creation of a unitary state was not viewed as an opportunity, two possibili-
ties were offered, that is to say, to create either a confederacy or a federation.
As is well known, the first constitution of the United States establishes a con-
federation. The member states are in the position of principals delegating some
tasks to their agent, the national government. To meet the requirements of a
confederacy, the mechanism of delegation must obey the principles of lateral
subsidiarity.

Establishing a framework for this lateral subsidiarity, the Articles enumerate
specific and supposedly limited tasks that are delegated to the central govern-
ment. The prerogatives of the latter are mostly directed to the provision of pure
public goods: National defense, open commerce, diplomacy are as many non
excludable and non rival goods. Such is the meaning of Article III: "The said
states hereby severally enter into a firm league of friendship with each other,
for their common defence, the security of their Liberties, and their mutual and
general welfare [...]." The exception is the duty of providing postal services
since the latter can be affected by crowding effects. Limits to the national
power are clearly delineated by Article II: "Each state retains its sovereignty,
freedom and independence, and every Power, Jurisdiction and right, which is
not by this confederation expressly delegated to the United States, in Congress
assembled."

Among the tasks delegated to the central government, one can notice that
there is listed the promotion of "mutual and general welfare," a vague and in-
determinate term that requires interpretation. One could consider that it gives
a wide basis to the power of the central government, being thus much more
of a federal type than of confederate inspiration. In fact, these words fit both
frameworks. It is thus not surprising that they are also used in the Consti-
tution of the U.S.A. (although they are abandoned by the 1861 Confederate
Constitution). The difference between confederation and federation, between
lateral and other forms of subsidiarity, rather lies in the extent and form of the
control of the central government. Lateral subsidiarity emphasizes bargaining
and varying agreements among the member states instead of a formalized and
more rigid covenant between the states and the central power. Two essential
features of confederation and lateral subsidiarity are indeed made explicit in
the Articles. Firstly, the central government is not granted with the autonomous
financial capacity to exert its power. Congress cannot raise funds directly from

citizens but must rely on the good will of the member states: "All charges of war, and all other expences that shall be incurred for the common defence or general welfare, and allowed by the united states in congress assembled, shall be defrayed out of a common treasury, which shall be supplied by the several states in proportion to the value of land within each state [...]. The taxes for paying that proportion shall be laid and levied by the authority and direction of the legislatures of the several states within the time agreed upon by the United States in Congress assembled" (Article VIII). Congress is thus the vehicle of lateral subsidiarity. Secondly, no compelling prerogatives are granted to the federal government. The absence of a central judicial power (there is no supreme court of justice) in the Articles clearly illustrates both the fact that collective decisions cannot but result from a bargaining process among the states and that central power is restricted by the very will of these states.

3.1.2.   The Limits of Lateral Subsidiarity   The principle of lateral subsidiarity established by the Articles of Confederation fails to provide cohesion to the fledgling union. The motivation of the confederacy is to withstand external forces that prove to be a military threat. States cannot by themselves solve the corresponding problem of size. Economies of scale, particularly in defense, must be reached but the instruments are largely ineffective. As was mentioned before, the conditions of an efficient bargaining are quite strong, and the lack of power of the federal government increases the difficulty. As is repeatedly stated by Hamilton for instance, a major flaw in the Articles is that "the government of the United States is destitute of energy" (Federalist Paper 15, 1987, p, 147). This problem has two dimensions. On the one hand, the federate level has "an indefinite discretion to make requisitions for men and money" (ibid., p. 148) but at the same time has "no authority to raise either by regulations extending to the individual citizens of America" (ibid., p. 148). To put it differently, the federal power cannot independently raise funds in order to produce or finance the provision of the public goods enumerated in the Articles. On the other hand, the "United States have no power to exact obedience, or punish disobedience to their resolutions" (ibid., p. 173). Hamilton thus stresses the lack of enforcement mechanisms in the delegation process.

For all that, and from the principle of lateral subsidiarity, it is not surprising that the federal level should have so few discretionary prerogatives. However, such a restriction can receive either a positive or a negative interpretation. The positive one conveys the idea that the control of the federal government is exercised by the states themselves: As principals, they should be able to contain and direct the actions of their agent (Holcombe, 1991, p. 314). On the negative side, one may emphasize the potential failure of a mechanism of control in which the states form the "ultimate instrumentality for executive action" (Ostrom, 1976, p. 6). Unfortunately, even if the experience of the Articles of

Confederation did not last for a long period, it provides a strong if not conclu-
sive case against the previous positive interpretation.

3.1.3.   A Confederation in Practice   Difficulties in enforcement are all
the trickier in that they relate to the provision of non excludable services. In
practice, the financing of the public goods provided by the Confederation is
through requisitions upon the states. The fledgling nation immediately faces
a typical problem of cost sharing. Confronted with the war effort, first on the
military then on debt, the confederacy is however granted no power to enforce
taxes (see Article VIII above). States themselves then have the responsibility
to raise funds on their own territory in order to pay their quota of the defense
effort. At this stage, we must remember that the Americans have successfully
resisted the British taxation policy. Attitude toward taxation during the confed-
erate period can be explained to a large extent by what happened in the colonial
period (Wallis, 1999, pp. 290-291). The arrangement provided by the Articles
proves cumbersome and, at the end of the day, quite ineffective. Congress re-
quires that a certain amount of money be collected to finance given projects.
The total cost of the requisition is then allocated among the states in propor-
tion to the shares specified by Article VIII. States cannot be forced to donate
the required amount, thereby letting the burden of public good financing fall
on the other contributors.

Dougherty and Cain (1997) give a simple but very intuitive explanation
(debated by Sobel, 1999) of the economic mechanism underlying the failure
of this process of cost allocation. Their model is adapted here in the following
way. Let $l \in \Re^+$ denote the amount of public good. The requisitioned level
is l_r. The numeraire is $x \in \Re^+$. Supplying a quantity l costs $x = c(l)$ to the
nation, where c is a cost function. Conversely, $l = c^{-1}(x) = c^{-1}(x_j + x_{-j})$
describes the amount of public good supplied when state s_j, $j = 1, \ldots, J$, con-
tributes x_j while x_{-j} designates the donations of the others. Since no enforce-
ment mechanism is available at the national level, states choose the amount
they contribute. If a state s_j keeps to its duty, then its share of the burden is
$c_j(l) = \alpha_j c(l)$. Parameters α_j, $j = 1, \ldots, J$ are such that $\sum_{j=1}^{J} \alpha_j = 1$ and
these proportions are constitutionally defined. For instance, Pennsylvania has
a share of 13.7% while Rhode Island bears 2.2% of the requisition base. Each
state benefits from the public good to an extent described by $b_j = b_j(l)$. Payoff
to state s_j is given by $b_j(l) - x_j = b_j(c^{-1}(x_j + x_{-j})) - x_j$. Assuming that all
the mathematical requirements on functions are met, a rational state chooses
its contribution x_j by calculating the unconstrained optimum of the program:

$$\max_{x_j} b_j(c^{-1}(x_j + x_{-j})) - x_j \tag{1}$$

It yields:

$$b'_j(l) = c'(l) \tag{2}$$

which obviously departs from the usual Lindahl-Bowen-Samuelson condition for the optimal provision of a public good. To illustrate it, Dougherty and Cain take the example of the army that is to be raised to counter Shays' rebellion in 1786. Congress requests $x^* = \$530,000$ in order to raise $l^* = 2,040$ soldiers, at a constant marginal cost. The proportion α_j for Maryland is 9.4%, as specified by the Articles. It would thus cost the state approximately \$50,000 to fulfil its duty. Virginia, the highest contributor with an α_j of 17.1%, intends to pay only part of its requisition, namely \$40,000, The other states do not seem to be willing to contribute. What will Maryland do? Paying the due share would mean bearing a burden of 50,000/90,000 = 56% instead of the constitutionally assigned 9.4%. History reveals that Maryland rationally does not contribute and that Congress must do with what Virginia accepts to donate.

Lateral subsidiarity requires an efficient incentive mechanism that could check the possible tendency of principals to avoid their duty. This may not happen with congested public goods where contributions would be set at the marginal cost of congestion, providing that such costs be shared knowledge. When quite pure public goods are to be provided, opportunism burdens the confederacy and the multi-principal setting appears to be largely inefficient, prompting reform toward a more authoritarian and centralist federalism.

3.2. The Assignment of Prerogatives among Agents in the US Federation

The shift from a confederacy to a federation requires a significant change in the agency contract. In particular, the constitutional assignment of the provision of public goods is now sustained by the corresponding power to tax and to control law enforcement (3.2.1). However, this new assignment may come under a great strain if the agents catch the opportunity to become more independent than is initially planned by the constitution (3.2.2).

3.2.1. States' Rights, Federalism and Subsidiarity The Convention of Philadelphia prepares the evolution from confederation to federation (3.2.1.1). It implies a move from bargaining among principals to competition between agents and from lateral to upwards subsidiarity. The corresponding reassignment of prerogatives results in a complete reshuffle of the agency relationship (3.2.1.2).

3.2.1.1.   From Confederation to Federation   The Articles of Confederation are very early criticized for their inability to constrain the member states. Concurrently, restrictions on the federal power would be too strong. As such, the criticism amounts to nothing more than a simple characterization of the properties of a confederation. Strategic behaviors during the bargaining process or at the stage of implementation of decisions can make the collective outcome drift away from Pareto efficiency. If that is the case, the passage to federation seems to be an unavoidable necessity. As Hamilton puts it, "there are material imperfections in our national system and [...] something is necessary to be done to rescue us from impeding anarchy" (Federalist Paper 15, 1987, p. 146), The solution then consists in modifying the agency relationship in order to evolve from lateral to upwards subsidiarity. But then a new constitution must be written: "After an unequivocal experience of the inefficiency of the subsisting federal government, you are called upon to deliberate on a new Constitution for the United States of America" (Federalist Paper 1, 1987, p. 1). However, the official aim of the Philadelphia convention is not so clear-cut since the latter is called "for the sole and express purpose of *revising* the Articles of Confederation and reporting to Congress and the several legislatures such alterations and provisions therein as shall; when agreed in congress and confirmed by the states, render the Federal Constitution adequate to the exigencies of government and the preservation of the Union" (Continental Congress, 21 February 1987, emphasis added). The convention brings about much more than a simple revision of the constitutional setting, it profoundly changes its nature from a system allowing and requiring bargaining among states to a constitution based on competition between agents. To reach this new goal, two steps are required.

Firstly, the principalship must be moved from the states to the citizens of the United States: "if it be possible at any rate to construct a federal government capable of regulating the common concerns and preserving the general tranquility [...] it must carry its agency to the persons of the citizens" (Federalist Paper 16, 1987, p. 154). This argument is embodied in the future federal constitution. The preamble displays the well known "We the people of America" whereas the introductory statements of the Articles of Confederation clearly consider the states as the principals. One can further notice that the source of power has been a matter of great debate during the Convention of Philadelphia (see also the Anti-Federalist Papers). Once the people is established as this source, there is no more justification to consider that the member states should bargain over the provision of public goods.

Secondly, the federal level will be granted with extended power. To use the words of Hamilton once again, if the national government is so destitute of energy under the Articles of Confederation, then the new constitution must provide it with "the means, and [...] methods, of executing the powers with

which it is entrusted, that are possessed and exercised by the governments of the particular states" (Federalist Paper 16, 1987, p. 154). The shift in principalship certainly cannot be expressed more clearly. It implies the capacity to raise funds and to control law enforcement which the previous constitutional arrangement so lacked.

Once accepted, the new setting replaces bargaining among principals with competition between agents. It also rests on another form of subsidiarity.

3.2.1.2.   The Assignment of Rights in the Federal Constitution   Upwards subsidiarity is here defined by a constitutional assignment of prerogatives to the national level. Three main features characterize the way in which these rights are granted respectively to the states and the federal government.

Firstly, the previous constitutional system based on bargaining between the states is abandoned. The former principals leave their role to the citizens. Concurrently, there is a shift to upwards subsidiarity since prerogatives are now constitutionally assigned to the federal level. As pointed out by Holcombe (1991), the national government created by the Articles is to report to the state legislatures, and is mostly run by them. Delegated tasks are quite limited, at least in their number if not in their scope. There is, moreover, no locus of final authority. The prerogatives of the national government are not stated once and for all. They are rather the fluctuating result of negotiations between states. Any breakdown of the bargaining may endanger the confederation. On the contrary, the 1789 constitution constructs the central level and carefully enumerates the rights and power of the federal government. In particular, in order to provide and finance public goods, Congress is granted with the capacity to raise taxes directly.

Secondly, if the federal power has extended prerogatives, constitutional safeguards must be created in order to permit and limit the discretionary power of the federal government. The constitution of a federal system requires "a solution to the problem of constitutional rule: How to devise a system of government where the rules of constitutional law could be enforced as against those who exercised the prerogatives of law making, law judging and law enforcing" (Ostrom, 1976, pp. 8-9). A corollary of this requirement is the relaxation of the rule of unanimity that prevailed under the Articles of Confederation. Under the new constitutional framework, states lose their power of veto on federal decisions, thus providing the national government with a real discretionary power. At the same time, the latter is checked by its own branches: The President has important prerogatives; Congress becomes bicameral, in place of the unicameral setting of the Confederation; A Supreme Court of Justice is created. The constitution organizes competition between these agents.

Thirdly, if as Hamilton puts it, "laws are a dead letter without courts to expound and define their true meaning" (Federalist Paper 22, 1987, p. 182), the role and place of the Supreme Court must be precisely investigated. It is

usually assumed that the Court is created to check the behavior of the other branches of government and to contribute to the balance of powers. More precisely, the Court is established in the position of enforcer of the constitutional assignment of rights. However, a second rationale for the existence of a Supreme Court of Justice can be found in Hamilton's words. Hamilton points out the "want of a judiciary power," arguing that "[i]f there is in each State a court of final jurisdiction, there may be as many different final determinations on the same point as there are courts. There are endless diversities in the opinions of men [...]. To avoid the confusion which would unavoidably result from the contradictory decisions of a number of independent judicatories, all nations have found it necessary to establish one court paramount to the rest, possessing a general superintendence and authorized to settle and declare in the last resort a uniform rule of civil justice" (Federalist Paper 22, 1987, p. 182). The famous decision *Fletcher v. Peck* (1810) canceling a law of the state of Georgia for its unconstitutionality is a direct application of the Hamiltonian principle. The words of John Marshall when he delivers the opinion of the court in this decision are quite clear: "Georgia cannot be viewed as a single, unconnected, sovereign power, on whose legislature no other restrictions are imposed than may be found in its own constitution [...] she is a member of the American union; and that union has a supremacy of which all acknowledge."

Beside the position of ultimate enforcer given to the Court, what is also worth noticing here concerns the nature of law: In economic terms, it is typically a public good for which there are spillover effects. This is one of the reasons why its provision must in Hamilton's opinion be federal rather than local. But will that induce a pervasive role of the Court, reinforcing possible failures of the competitive mechanism that is supposed to bind the agents into their prerogatives?

3.2.2.   The Trend toward Centralization   The way upwards subsidiarity is organized contains dangers present but hidden at the original constitutional moment. The assignment of rights is not in practice as strict as the enumeration in the constitution claims it to be. The door is open for more centralization (3.2.2.1). The virtues of federalism are not as straightforward as it may seem at first glance. Never really considered as an opportunity, the unitary solution is not necessarily always inferior to the federal one (3.2.2.2).

3.2.2.1.   How Far is the Assignment of Rights Unstable?   The assignment of prerogatives to the federal government is supposedly justified on the ground of economic efficiency. Negotiations on the range and financing of goods that should be provided is replaced by constitutional assignments. In this respect, there are two opposed ways of reaching an agreement on these

assignments. The first one is an ongoing process of bargaining while the second one rests on an a priori decision. Then rights granted ex ante to the federal government of providing public goods must have an objective basis, otherwise "assignment ceases to work as an independent federalist institution" (Inman and Rubinfeld, 1997b, p. 95). Indeed, if the nature of the goods is such that "a given activity's spillover becomes open to dispute, then the application of the assignment principle becomes political" (ibid.).

Discussions on the eighth section of the first article of the Constitution show how difficult it is to combine, on the one side, the right to raise funds to finance public goods with, on the other side, the possibility to contain the exercise of this right to the sole provision of pure or at least "federal" public goods. The obvious problem is the definition of what these goods are. For instance, during the first Congress, there are many discussions about the federal financial support for the building of a lighthouse, namely on Cape Henry in Virginia.

The situation is all the more intricate since the constitution fails to make clear the extent of state powers. The Tenth Amendment, added in 1791, attempts to remedy this situation by reserving to the states the powers not delegated to the national level. What is here labeled as federal dualism soon gives rise to the states' rights doctrine. The latter will not abandon their principalship without fighting. One of the main episodes is the South Carolina nullification ordinance of 1832, contesting the capacity of Congress to impose "laws laying duties and imposts on the importation of foreign commodities" and announcing the Secession of 1860–1861.

One could argue that the role of a third party is to settle disputes, here between the states and the federal government, and thus to allow the federation to work. Placed in the position of an ultimate enforcer, the Supreme Court logically plays this role; It also benefits from the incompleteness of the constitutional contract concerning its prerogatives. At first, no basis exists to ground a common law for all the United States of America. Therefore, the judges of the Court soon take the opportunity of this constitutional void to assume an increasing power (Josselin and Marciano, 2000) and competition between agents soon loses its fine edge. For instance, Marshall reiterates the position he made clear in *Fletcher v. Peck* with cases like *Gibbons v. Ogden* (1824). After the civil war, the trend towards centralization meets little resistance. But is it necessarily the most efficient solution?

3.2.2.2.   Provision of Law: Is Federalism always Efficient?   The model of Rose-Ackerman (1981) provides a very fine demonstration that things are not as simple as they seem to be, at least in the economics of constitutions. It considers a federate republic with two layers, the national one and the state level. State legislative choices can always be pre-empted by the federal government, which is consistent with the usual power of the Supreme

Court. The American case appears as a watermark but the purpose of the model is more general of course.

The setting described here is very simple (Rose-Ackerman provides more elaborate schemes). Immobile individuals vote on single-issue legislative choices. Democracy is direct and people cannot vote with their feet whereas capital is mobile. More specifically, the object of the vote is the passing of a new law on a given subject. The initial situation, absence of law, characterises the status quo $\bar{l}$. For example, casino gambling is illegal within the state. This status quo completely defines the initial situation in a unitary system. Legislation l that permits casino gambling and levies a tax on earnings is passed if and only if:

$$N(l \succ \bar{l}) > N(\bar{l} \succ l) \tag{3}$$

where N is the number of voters who prefer one situation to the other and $\succ$ denotes an ordering of weak preferences.

In a federal system, however, some states may have declared casino gambling illegal ($\bar{l}$) while others may have already passed a law l accepting and taxing it. Again let s_j, $j = 1, \ldots, J$ denote the states in the federal system. If a state s_j accepts gambling then $s_j = s_j(l)$. If there is no such state law in s_j then $s_j = s_j(\bar{l})$. The status quo in a federal system is a vector SQ such that $SQ = (s_1, \ldots, s_j, \ldots, s_J)$ with $s_j(l) = l(SQ)$ describing states in which the law has already been passed and $s_j(\bar{l}) = \bar{l}(SQ)$ denoting states in which there is no such law at the moment. Individuals live in $\bar{l}$-*type* states or in l-*type* states. Legislation l passes at the national level if and only if:

$$N_l(l \succ l(SQ)) + N_{\bar{l}}(l \succ \bar{l}(SQ)) > N_l(l(SQ) \succ l) + N_{\bar{l}}(\bar{l}(SQ) \succ l) \tag{4}$$

Beginning from the left of the equation, the first term gives the number of individuals living in l-*type* states who prefer national legislation l to the existing system (they thus favour the extension of their own state law to the federal level). Voting in a similar way are the individuals from $\bar{l}$-*type* states who also favour federal law without having it at the moment in their own state. The last two terms of equation (4) describe the opposite opinions.

Equations (3) and (4) need not bring the same result. In order to further compare the unitary and federal systems, the following notations are introduced. $L(SQ)$ (respectively $\bar{L}(SQ)$) is the set of individuals i living in states where the law is already in place (respectively where there is no such law). Within $L(SQ)$ there are four categories of voter preferences. $L_1 = \{i \in L(SQ), l \succ \bar{l}$ and $l \succ l(SQ)\}$ describes the set of individuals who in a unitary system would prefer the passing of the law and who would favour its extension at the federal level.

<div style="text-align:center">

TABLE 1.
Voter's Position in Unitary and Federal Systems

</div>

Federal or unitary	$l \succ l(SQ)$	$l(SQ) \succ l$	$l \succ \bar{l}(SQ)$	$\bar{l}(SQ) \succ l$
$l \succ \bar{l}$	L_1	L_3	$\bar{L}_1$	$\bar{L}_3$
$\bar{l} \succ l$	L_2	L_4	$\bar{L}_2$	$\bar{L}_4$

For people in $L_2 = \{i \in L(SQ), \bar{l} \succ l \text{ and } l \succ l(SQ)\}$, the status quo is prefer-able in a unitary system while they would vote for a federal law. The opposite case is $L_3 = \{i \in L(SQ), l \succ \bar{l} \text{ and } l(SQ) \succ l\}$ where individuals prefer the passing of the law at the unitary level but would vote against the federal exten-sion of state legislation. Finally, the set $L_4 = \{i \in L(SQ), \bar{l} \succ l \text{ and } l(SQ) \succ l\}$ comprises those voters who prefer the status quo in both a unitary and a fed-eral system. Sets $\bar{L}_1, \bar{L}_2, \bar{L}_3, \bar{L}_4$ are similarly defined. The possible cases are summed up in Table 1.

For instance, voters in $\bar{L}_3$ oppose federal law but favour unitary legislation. In a unitary system, the number of people who favour the law is:

$$N(l \succ \bar{l}) = N(L_1 \cup L_3 \cup \bar{L}_1 \cup \bar{L}_3) \tag{5}$$

whereas in a federal system it amounts to:

$$N_l(l \succ l(SQ)) + N_{\bar{l}}(l \succ \bar{l}(SQ)) = N_l(L_1 \cup L_2) + N_{\bar{l}}(\bar{L}_1 \cup \bar{L}_2) \tag{6}$$

Does the unitary system systematically favour the passing of centralized legis-lation? The vote for the law in this setting is greater than the vote in a federal system if, using equations (5) and (6):

$$N(L_3 \cup \bar{L}_3) > N(L_2 \cup \bar{L}_2) \tag{7}$$

The left term of the equation describes the number of people who prefer the law in a unitary system but who benefit from the absence of federal law in a federal system. The right term counts individuals who prefer a status quo in a unitary system but who favour federal legislation. However, since inequality:

$$N(L_3 \cup \bar{L}_3) < N(L_2 \cup \bar{L}_2) \tag{8}$$

may hold instead of (7), federalism "matters," to use the phrase of Rose-Ackerman (1981, p. 156). Depending on strategic behavior of citizens, fed-eralism may even favour centralization more than a unitary system: Whenever equation (8) holds, the law gets more votes in the federal context.

The trend towards centralization logically raises the question of devolution. In this respect, a 1985 ruling of the Court *(Garcia v. San Antonio Metropolitan Transit Authority)* denies the states the right to seek redress from the Court for federal encroachment of their powers under the Tenth Amendment. Though the Court has since seemed to become more supportive of states' rights (Kincaid, 1998), centralized federalism still remains the dominant feature.

3.3.   Conclusion of Section 3

Contests as to the principalship are not simply theoretical matters. In this process, the U.S.A. hardly avoids a breakup of the nation. The "perpetual" and "indestructible" union is soon endangered by the threat of secession that is explicitly stated in the South Carolina Ordinance of Nullification. History shows how contention grows to war after this fatal breach is open. The long experience of the American federalism may provide as many lessons for Europe. A clear definition of who the principal is, a clear assignment of prerogatives among agents are all necessary conditions for an efficient and stable organization of government. Subsidiarity, already a pivotal though implicit concept in a national context, becomes a founding principle when it comes to federating the nations of Europe.

4.   FEDERATING NATIONS: SUBSIDIARITY IN INTERNATIONAL CONTEXT

Federating nations necessarily differs from federating a nation. In the latter case, the feeling of membership in a same entity preexists and underlies the process even if this may not go without tragic periods, as is illustrated by the history of the U.S.A. The former case, since it consists in the gathering of independent nations into a single political body, is rather different and, in some respects, more complex. Federating nations may require the harmonization of different political systems and legal orders. There may already exist federations as well as unitary states. Countries of common law tradition may have to join with statute law nations. The requirements constraining the political entrepreneurs who lead the process of federating nations are thus likely to be tighter than those that frame the federation of a nation. A manifest instance of the former type of situation is given by the ongoing process of integration in Europe. In this section, focus will be put on the way subsidiarity is organized in the European Union. In order to analyze it, one could be tempted to parallel the situation in Europe with the American constitutional history (Boom, 1995; Vibert, 1995; Josselin and Marciano, 2000). However, beside obvious and sometime striking similarities, there are also substantial differences among which the major one is the fact that Europe still has difficulties in choosing between a confederate and a federate institutional structure (Koslowsky, 1999).

Up to now, it has always been staying half-way between the international legal order of a confederacy, and the constitutional order of a federation (4.1). As a consequence, the assignment of prerogatives remains far from stable (4.2).

4.1. The Economics of an International Legal Order: The European example

At the end of World War II, some political entrepreneurs (Jean Monnet or Robert Schuman, for instance) are led to consider the possible melting of the European nations into a unified political entity. However, the not so distant conflict makes it impossible to build it at once. Europe is thus created as an international organization ruled by international public law. To study it, the approach adopted here departs from the usual economic analysis of international organizations (for a survey and a comprehensive bibliography, see Frey, 1997a) in that we apply the agency theory to understand the way powers and prerogatives are assigned in this international order. As such, Europe displays some of the characteristics of a confederacy (4.1.1)—although the word has rarely been used as a label for the European institutional structure—as well as features of a federation, making it a hybrid system (4.1.2).

4.1.1. Attributes of a Confederation The prerogatives granted respectively to the member states and to the supra-national level depend on the legal nature of the treaties upon which the European Union is progressively built. Among others, Schilling (1996) or Weiler and Halpern (1996) consider that Europe is created as an international legal order, thus conveying the idea that the first treaties (European Coal and Steel Community or ECSC, European Economic Community or EEC, European Atomic Energy Community or EURATOM) are not conceived as a constitution designed by an autonomous and original constituent power. Admittedly, other legal scholars interpret ratification procedures by the national legislatures as constituent acts. In this perspective, the treaties could be seen as forming the body of a constitution, and thereby European citizens could be considered as the source of power; National legislatures would be the vehicle of this sovereignty. In economic terms, the individuals would be the principals, the institutions of the Community would be their agents and by no means those of the states. However, nothing indicates that ratification does amount to exercising a real constituent power. It is rather intergovernmental conferences that have finalized documents drafted mainly through diplomatic bargaining. An international organization is indeed created by the first treaties, "but with no measure of independence or power to eradicate its subordination to its States parents and its subjection to the classical laws governing the States' treaty relations" (Weiler and Halpern, 1996,

p. 8). Thus, in the international legal order, the member states are in the position of principals delegating some tasks to their agents, the European supranational institutions. They remain the *"Herren der Verträge"* (the masters of the covenants), to use the German phrase.

An opinion put forward in 1957 confirms and makes this interpretation clearer: "[t]he Treaty [ECSC treaty] is based upon delegation, with the consent of the Member States, of sovereignty to supranational institutions for a strictly defined purpose [...] The legal principle underlying the Treaty is a principle of limited authority" (joint cases 7/56 and 3-7/57, *Dineke Algera et al. v. Common Assembly of the European Steel and Coal Community,* 1957, E.C.R 69, opinion of Mr. Advocate General Lagrange, 69, p. 82). As will be repeatedly stated in Community decisions, the European Community is granted with limited competencies, the extent of which is determined by diplomatic bargaining, treaty after treaty. This is the first requirement of lateral subsidiarity displayed by the delegation process in the original treaties.

A second principle of lateral subsidiarity is of major importance: Member states must have a veto power. Such is the case in Europe from its earlier stages to the 1986 Luxembourg compromise. The decision making process is at that time based on the rule of unanimity. States must bargain to reach an agreement. This requirement is quite effective. For instance, it causes the failure of a treaty establishing a European Community of Defense. Drafted in 1952, the treaty is ratified by only four out of six countries (Italy and France reject it). Beyond strategic and international relations, this is once again an illustration of the difficulty of providing a public good when the enforcement mechanism is weak or even absent. The analysis of Dougherty and Cain (1997) finds here another application.

The Luxembourg compromise is the first departure from the confederate setting. It brings about incentives to shift from bargain federalism to a more centralized vision. However, this shift remains partial and the move toward a federation quite hesitant, which blurs the original design of the founding fathers of Europe.

4.1.2.   Attributes of a Federation   The European legal order undoubtedly possesses characteristics of an international organization, thus being based on principles of lateral subsidiarity. Nevertheless, from its very inception, it contains the seeds of a federation. In other words, the European community is created as a confederation with some already federate features. In particular, subsidiarity is both of a lateral and of an upwards type.

Upwards subsidiarity shows through in many instances. The European Coal and Steel Community treaty grants a high authority with compelling prerogatives. Article 189 of the Treaty of Rome creates a category of norms called

regulations "that do not require national implementing measures but are binding on the states and their citizens as soon as they enter into force" (Mancini, 1991, p. 181). Another significant point is the creation of a European Court of Justice. As we have seen previously, the existence of a supreme court can be justified by using two arguments. The court is first a means to check the other components of government. Secondly, law is a public good whose spillover effects must be controlled at the highest level, even at the cost of possibly more centralization than in a unitary framework. However, these arguments are somewhat misplaced when they are put forward in a strictly confederate framework. A court of justice, within the setting of an international legal order, can be viewed either as a third party enforcer, in which case it solves international legal conflicts by using international public law, or as the court of a federal system. Obviously, subsidiarity does not receive the same interpretation in the two systems. The way in which the European Community has evolved seems to suggest that the second path is being followed, breeding a constitutional order and its subsequent assignment of prerogatives.

4.2.   The Assignment of Prerogatives in the European Constitutional Order

The European Community soon embodies the characteristics of a constitutional order (4.2.1) in which the assignment of prerogatives may evince some instability (4.2.2).

4.2.1.   The European Constitutional Order   A shift from an international to a constitutional legal order necessarily means a change in the agency relationship (4.2.1.1). The constitutional framework then defines the way powers and prerogatives are assigned. A related acceptation of subsidiarity is embedded in this transformation. The new legal order nonetheless displays hybrid and complex characteristics (4.2.1.2).

4.2.1.1.   From an International to a Constitutional order   The set of original treaties (ECSC, EEC, EURATOM) has rapidly been interpreted by the European Court of Justice, according to what were supposedly the intentions of the framers of the European Community. The major decision of the ECJ in this perspective is the prominent case *Van Gend & Loos v. Nederlandse Administratie der Berlastingen* (case 26/62, 1963, E.C.R 1). The court points out what can be viewed as a central characteristic of the Treaty of Rome: "This Treaty is more than an agreement which merely creates mutual obligations between the contracting States." It would then belong to a legal category quite different from the one—international public law—initially conceived during the early stages of the Community. The Court thus emphasizes the need for a

European constitution, the treaties forming its original components. The major activity of the Court is soon qualified as "judicial activism" since it consistently aims at this constitutionalization (see for instance Burley and Mattli, 1993; Garret, 1992; Stone, Sweet, and Caporaso, 1998). Progressively using the legal doctrines of supremacy, direct effect, implied powers and preemption (Stein, 1981; Mancini, 1991), the Court masters this process, crowned by the definition of the Treaty of Rome as the European "basic constitutional charter" (case 294/83, *Parti écologiste "Les Verts" v. European Parliament,* 1986, E.C.R. 1339, p. 1365).

In economic terms, the move from an international legal perspective to a constitutional framework necessarily changes the prerogatives within the agency contract. While from the perspective of international law a supranational government remains subordinated to its creators, the reverse occurs in a constitutional order. By the means of a constitution, the constituent units subordinate themselves to their creation. This change affects the basic nature of the agency relationship: "whereas the subjects of a treaty (or a treaty-based organization) are the states composing it, the subjects of, say, a federal constitutional order are not only its constituent states, but also its common citizenry. This difference is thought to create a different level of legitimacy for the constitutional order, one where its legitimacy does not come only from the consent of sovereign states but from the broader and more direct consent of the citizens of those constituent units. Typically, the international organization is governed by international law and the constitutional order by its own municipal law" (Weiler and Halpern, 1996, p. 10). In other words, the evolution from an international legal order to a constitutional one changes the principalship. It also creates a "municipal law" endogenous to the contract, whereas the former agency relation was set in a largely exogenous juridical framework, that of international public law.

As was mentioned before, this transformation has its origin in the *Van Gend en Loos* case. The Court states that the Community is "a new legal order of international law [...] the subjects of which comprise not only the Member States, but also their nationals," and adds that "the task assigned to the Court under Article 177, the object of which is to secure uniform interpretation of the Treaty by the national courts and tribunals, confirms that the States have acknowledged that Community law has authority which can be invoked by their nationals before their courts or tribunals." Here, the point is not only that the Court reminds the member states, the principals, that their authority is limited but mainly that there is a new legal order of constitutional nature whose subjects are not only the member states, but also the citizens.

The new setting resulting from *Van Gend en Loos* is intricate. A real and effective move from a confederation to a federation would have required to straightforwardly replace the principalship of the member states by that of the

European citizens. Such was the way the Articles of Confederation were abandoned for a constitution in which the people of America was the principal, even if advocates of the states' rights doctrine rejected the notion sometimes vehemently. On the old continent, the European Court of Justice only partially establishes a new agency relationship by adding a second source of principalship. The co-existence of these two sources, and the ensuing multi-principal setting, have important consequences on the assignment of rights between the states and the supranational power.

4.2.1.2.   **The European "Constitution" and the Assignment of Powers** The way rights are assigned between the states and the European Community necessarily results from the existence of the two levels of principalship. Firstly, citizens can sue their own state through the European legal channel. It clearly indicates a decrease in state prerogatives. Secondly, bargaining among principals of the same level—a confederate principle—is largely abandoned. These two related points are linked by the conception of law underlying the *Van Gend en Loos* case. The argumentation put forward by the Court is that law is a public good with spillover effects and as such it must be provided at the central level rather than at the local or national levels. Thus, the existence of a "common legal order" firstly means that law has to apply directly and uniformly within each country member of the Union, and secondly that state legislatures and judiciaries are only part of this legal order. No local or national power should interfere with the application of law over all the European countries. The process of harmonization, particularly because of the doctrine of supremacy of Community law, is a perfect illustration of such a conception (Marciano and Josselin, 2002). This perspective typically corresponds to the definition of upwards subsidiarity. However, the rights and prerogatives are not assigned to the central level by a formal constitution, but rather by a judicial constitutionalization of treaties.

In the absence of any constituent assembly, the European Court of Justice largely proceeds to the assignment of powers. Indeed, the constitutionalization of treaties leads to a "mutation of competencies" (Weiler, 1991) in the European Community. More precisely, the rights granted to the Court progressively increase. As a result, the constraint conveyed by the principle of enumerated powers seems to have disappeared (Weiler, 1991), and "[t]here is simply no nucleus of sovereignty that the member states can invoke, as such, against the Community" (Lenaerts, 1990, p. 220). To put it differently, the shift of principalship to the citizens not only results in the decrease of power of the original principals, the member states; It also has the consequence of increasing the power of the European Court of Justice. The latter establishes itself in the position of a final authority, required to resolve conflicts between principals.

In general terms, upwards subsidiarity increases the prerogatives granted to the federal level. Constitutional safeguards must then be created to limit its

discretionary power: A federal constitution requires a system of checks and balances such as the one introduced in the U.S. 1789 constitution. In this perspective, the role and functions of the European parliament are of a particular importance. Created only as a consultative structure, the elected assembly can now check and balance the power of the Court, mainly through the co-decision procedure introduced in the Maastricht treaty. Another constitutional safeguard is the "subsidiarity" principle as it is expressed in this treaty: The Community is to take action in areas which do not fall within its exclusive competence only if the objectives of the proposed action cannot be achieved by the member states. Thus, "[t]he concept of Community law supremacy and the subsidiarity principle are expressions of contradictory evolutions of the relationship between Community and national laws" (Schilling, 1995, p. 13). This subsidiarity principle may be a way to reduce the discretionary power of the European Community. As such it has been considered as the word that saved Maastricht (Cass, 1992) and viewed as a principle likely to "safeguard the Member States national identity and preserve their competencies" (German Federal Constitutional Court, Maastricht decision). It has been confirmed by the Amsterdam Treaty (Article 5).

The evolution of the European institutions takes the form of a transformation of the original treaties in a constitutional setting that has never been conceived by a constituent assembly, but rather through the judicial activism of the Court. The corresponding modification of the agency relationship informs us about the respective prerogatives of the member states and the central, supranational institutions. This constitutional evolution and the related assignment of rights has to a large extent been directed by the European Court of Justice, a highly powerful agent.

4.2.2.   An unstable assignment of powers   As is the case in the U.S.A., the constitutional framework developed in the European Union largely rests on upwards subsidiarity. It is interpreted both in the Treaties and by the European Court of Justice, which leads to a quite specific assignment of prerogatives with no definite boundary (4.2.2.1). Is this assignment the most efficient way of controlling the agents (4.2.2.2)? Is centralized federalism the only way (4.2.2.3)?

4.2.2.1.   How far is the assignment of rights unstable?   If the principle of subsidiarity is as much a legal rule as it is a political one, then the question of its justiciability lies in the hands of the European Court of Justice. The latter thus has some effective control over the assignment of rights. It is all the more the case since the possible conflict between the implied-powers doctrine and the doctrine of enumerated powers remains pending (Kirchner, 1997). At the same time, the Treaty must find some balance between the two levels of principalship. The second section of Article 3B explains that "the Community

shall take action, in accordance with the principle of subsidiarity, only if and in so far as the objectives of the proposed action cannot be sufficiently achieved by the Member States and can therefore, by reason of the scale or effects of the proposed action, be better achieved by the Community." This shows how the principle of subsidiarity is focussed on the relationship between Community and member-states. The "European people" is given a less clear place, although the preamble of the Maastricht Treaty states that "decisions are taken as closely as possible to the citizen." The concurrence of these two statements acknowledges the existence of the two levels of principalship. It also exemplifies the difficulty of federating nations. Whereas in the U.S.A., conflicts have often arisen from the question of states' rights, the people in Europe may be granted a quite elusive role. In this respect, the Court remains in its position of final arbitrator. It increasingly emphasizes people's rights, particularly when member states fail to comply with European directives. An example of it is the 1991 *Francovich case* (*Andrea Francovich and Danila Bonifaci and others v. Italian Republic,* joined cases C-6/90 and C-9/90, E.C.R 1991) by which individuals can claim compensation before a national judge after suffering damage resulting from a lack of implementation of a European directive at the state level.

Subsidiarity in the new European order is initially conceived as a way of dealing with environmental issues. If one considers the limits of the Tiebout model, environment is one of the major fields in which rights and prerogatives are usually assigned to the supranational level. The rationale for this is the possible race-to-the-bottom implied by prisoners' dilemma competition and transboundary spillover effects. On the other hand, the risk of such a centralization is the possible inefficiency of the central institutions. It reinforces the need for a strict control of the agents.

4.2.2.2.   Collusion or competition between agents   The prerogatives of the European parliament shift from consultation to actual power in decision making once the co-decision procedure is introduced. Is it possible to consider that this is a genuine way to check the behavior of the other agents and to encourage competition between them? Just like in the U.S.A. where similar mechanisms progressively lose their fine edge, co-decision rather results in an increase in the power of the Court (Cooter and Ginsburg, 1999). Using a game-theoretic analysis, Cooter and Drexl (1994) show that whenever the European parliament gets more power, the discretionary prerogatives of the Court increase, each time to the detriment of the Commission. Since the latter represents the principals, there is a shift of power from the member states to their agents. This process of democratization should enhance the second source of principalship, namely the citizens. However, this is done through a kind of bargaining between the European parliament and the Court, and not through a

constitutional moment. This logrolling process aims at a reciprocal increase in their respective prerogatives. While rulings of the Court promote parliament's rights, diplomatic negotiations before the Maastricht Treaty lead to a request of parliament according to which the Court should be given a right to act against these member states that would fail to respect its judgments (Corbett, 1994). Once again, all this is achieved not by the means of a constituent assembly, but rather through diplomatic bargaining.

4.2.2.3.   Why and which federalism for Europe?   It is not at all sure whether all the opportunities offered by the concept of subsidiarity have been explored in practice (see an application to environmental issues in Backhaus, 1997b). The trend, or at least the willingness, to come to more devolution in the U.S.A., could raise similar debates in Europe. Downwards subsidiarity also has an institutional translation with the FOCJs and this solution could be explored more thoroughly. Moreover, this new functional federalism can be associated with a growing constitutional liability of regions, as they are defined for instance by Drèze (1993) or Josselin and Marciano (1999c). This functional and spatial competition can also be extended to the legal domain. In this respect, Carney (1997) shows that company law directives in Europe are designed to provide harmonization. The Treaty of Rome establishes it with Articles 54 and 58. The idea is to facilitate the creation of a common market. This is quite logical when the goal is to federate nations. It may however protect rent extraction by existing interest groups. This rent-seeking would otherwise be weakened in a setting of competition on the market for corporate charters. The argument must of course be reversed in the American case. Federating a nation resulted there "in the early creation of a common market […] which created corporate mobility in choosing a chartering state that could not readily be thwarted by any single jurisdiction" (Carney, 1997, p. 329). As such, the American system seems to produce corporate laws providing less regulations than the European one does.

   But why should we take federalism for granted? The model of Rose-Ackerman (1981) shows that the issues at stake are not so readily solved by choosing a federate structure. After all, some states in Europe display a unitary structure with some kind of success, so why not a unitary European state? Two arguments may corroborate this view. The first one concerns the citizens as a first source of principalship. They may want to strategically take advantage of spillovers in the provision of public goods. In some circumstances, this may lead to more centralization in a federation than in a unitary structure, as is demonstrated by Rose-Ackerman. The second argument also borrows from the American case. The principalship of the member states may progressively be contested and since Europe is still half-way between confederation and federation, it may suffer from the same evils as the U.S.A. did with the states' rights

doctrine. When centralized federalism was not firmly established, the lack of a clear boundary between federal and state powers gave rise to such troubles that were described in the previous section. This may not happen in Europe, thus realizing the ambition of its founding fathers.

4.3. Conclusion of Section 4

The European Union is a quite young institution and one cannot expect it to be entirely stable. Federating nations implies that principalship progressively changes: At first, member states are the principals, then the people acquire this status. Unlike the American case, this is done not through a "constitutional moment," like the Philadelphia convention, but rather through the judicial "activism" of the European Court of Justice. If federalism is to be preferred to the unitary solution, then many paths have yet to be explored. In particular, downwards subsidiarity would enhance competition among functional and regional institutions. Citizens, as principals, could find there an efficient way of delegating their inherent rights and of controlling their use by the agents.

5. CONCLUSION: FOUNDATIONAL CONCERNS

If we accept methodological individualism, then the status of individuals, regarding public decisions, should be at the heart of the constitutional debate. In economic terms, they should be considered as the ultimate or primary principals. The problem is then to devise an adequate agency relationship, capable of reaching the assigned goals. Foundational concerns emerge when the agents or other levels of principalship seem to acquire too much power. The problem of this agency is fundamentally different from that in labour economics for instance, where the worker, as an agent, must certainly be safeguarded. In the constitutional game, agents must remain instruments of the people. This instrumentality is at risk whenever the servant threatens to become the master of his master, to use the Hegelian metaphor. It necessarily raises the question of individual and governmental liabilities. They seem to be usually defined by judicial means rather than by constitutional moments. Furthermore, they are not necessarily better assured in a federal context than in a unitary one.

Concern does not necessarily convey pessimism. However necessary the control of government, there remains an economic rationale for confidence in the contracting parties: There should be no a priori prejudice or assumption as to the nature of the principal and of the agents in order to avoid crowding out of civic virtues (Frey, 1997b). Humean sympathy in the agency contract would of course solve many incentives problems. Delegation of power in a democracy nevertheless remains a genuine principal-agent question: "[t]he most efficient government is not the most orderly looking government but the government that comes closest to carrying the wishes of its masters" (Tullock, 1969, p. 29).

REFERENCES

Aghion, Philippe and Tirole, Jean. (1997). "Formal and Real Authority in Organizations." *Journal of Political Economy* 105 (1): 1-29.

Alesina, Alberto and Spolaore, Enrico. (1997). "On the Number and Size of Nations." *Quarterly Journal of Economics* 112 (4): 1027-1056.

Anderson, Terry L. and Hill, Peter J. (1977), "An American Experiment in Anarcho-capitalism: The Not So Wild, Wild West." *Journal of Libertarian Studies* 3 (1): 9-29.

Anderson, Terry L. and Hill, Peter J, (1986). "Constraining the Transfer Society: Constitutional and Moral Dimensions." *Cato Journal* 6 (1): 317-339.

Anti-Federalist Papers and Constitutional Convention Debates. (1986). Ralph Ketcham, ed. New York: Mentor.

Aranson, Peter. H. (1988). "Procedural and Substantive Constitutional Protection of Economic Liberties." In *Public Choice and Constitutional Economics,* Gwartney, James D. and Wagner, Richard E., eds., pp. 285-313. Greenwich: London: JAI Press.

Backhaus, Jürgen G. (1997a). "Christian Wolff on Subsidiarity, the Division of Labor and Social Welfare." *European Journal of Law and Economics* 4 (2-3): 129-146.

Backhaus, Jürgen G. (1997b). "Subsidiarity and Ecologically Based Taxation: A European Constitutional Perspective." *Public Choice* 90 (1-4): 281-310.

Backhaus, Jürgen G. (1999). "Subsidiarity." In *The Elgar Companion to Law and Economics,* Backhaus, Jürgen G., ed., pp. 136-143. Aldershot: Edward Elgar.

Backhaus, Jürgen G. and Wagner, Richard E. (1987). "The Cameralists: A Public Choice Perspective." *Public Choice* 53 (1): 3-20.

Bewley, Truman F. (1981). "A Critique of Tiebout's Theory of Local Public Expenditures." *Econometrica* 49 (3): 713-740.

Bish, Robert. (1988). "Federalism: A Market Economics Perspective." In *Public choice and Constitutional Economics,* Gwartney, James D. and Wagner, Richard E., eds., pp. 351-368. London: JAI Press.

Blum, Ulrich and Dudley, Leonard. (1991). "A Spatial Model of the State." *Journal of Institutional and Theoretical Economics* 147 (2): 312-336.

Boadway, Robin and Flatters, Franck. (1982). "Efficiency and Equalization Payments in a Federal System of Governments. A Synthesis and Extension of Recent Results." *Canadian Journal of Economics* 15 (4): 613-633.

Bolton, Patrick and Roland, Gérard. (1997). "The Break-up of Nations: A Political Economy Analysis." *Quarterly Journal of Economics* 112(4): 1057-1090.

Boom, Steve J. (1995). "The European Union after the Maastricht decision: Is Germany the "Virginia of Europe"?" *Jean Monnet Working Paper.* Harvard Law School.

Breton, Albert and Scott, Anthony. (1978). *The Economic Constitution of Federal States.* Toronto: University of Toronto Press.

Buchanan, James M. (1965). "An Economic Theory of Clubs." *Economica* 32 (1): 1-14.

Buchanan, James M. (1990). "The Domain of Constitutional Economics." *Constitutional Political Economy* 1 (1): 1-18.

Buchanan, James M. and Goetz, C. J. (1972). "Efficiency Limits of Fiscal Mobility: An Assessment of the Tiebout Model." *Journal of Public Economics* 1 (1): 25-43.

Buchanan, James M. and Tullock, Gordon. (1962). *The Calculus of Consent: Logical Foundations of Constitutional Democracy.* Ann Arbor, Michigan: The University of Michigan Press.

Burley Anne-Marie and Mattli, Walter. (1993). "Europe before the Court: A Political Theory of Legal Integration." *International Organization* 47 (1): 41-76.

Carney, William J. (1997). "The Political Economy of Competition for Corporate Charters." *The Journal of Legal Studies* 26 (1): 303-329.

Casella, Alessandra and Frey, Bruno S. (1992). "Federalism and Clubs: Towards an Economic Theory of Overlapping Political Jurisdictions." *European Economic Review* 36 (2-3): 639-646.

Cass, D. (1992). "The Word that Saved Maastricht? The Principle of Subsidiarity and the Division of Powers within the European Community." *Common Market Law Review*.

Cooter, Robert. (2000). *The Strategic Constitution.* Princeton: Princeton University Press.

Cooter, Robert and Drexl, Josef. (1994). "The Logic of Power in the Emerging European Constitution: Game Theory and the Division of Powers." *International Review of Law and Economics* 14 (3): 307-326.

Cooter, Robert and Ginsburg, Tom. (1999). "Division of Powers in the European Constitution." In *The New Palgrave Dictionary of Economics and the Law,* Newman, Peter, ed.

Corbett, Robert. (1994). "Representing the People." In *Maastricht and beyond. Building the European Union,* Duff, Andrew, Pinder, John, and Price, Roy, eds., pp. 207-228. London: Routledge.

Donahue, John D. (1997). "Tiebout? Or Not Tiebout? The Market Metaphor and America's Devolution Debate." *Journal of Economic Perspectives* 11 (4): 73-81.

Dougherty, Keith L. and Cain, Michael J. G. (1997). "Marginal Cost Sharing and the Articles of Confederation." *Public Choice* 90 (1-4): 201-213.

Dreschler, Wolfgang. (1999). "Christian Wolff (1679–1754)." In *The Elgar Companion to Law and Economics,* Backhaus, Jürgen G., ed., pp. 527-531. Aldershot: Edward Elgar.

Drèze, Jacques. (1993). "Regions of Europe: A Feasible Status, to be Discussed." *Economic Policy* 8 (17): 265-287.

Federalist Papers (1987). Isaac Kramnick, ed. New York: Penguin Books.

Frey, Bruno S. (1990). "Institutions Matter: The Comparative Analysis of Institutions," *European Economic Review* 34 (2-3): 443-449.

Frey, Bruno S. (1996). "A Directly Democratic and Federal Europe." *Constitutional Political Economy* 7 (4): 267-279.

Frey, Bruno S. (1997a). "The Public Choice of International Organizations." In *Perspectives on Public Choice,* Mueller, Dennis C., ed., pp. 106-123. Cambridge: Cambridge University Press.

Frey, Bruno S. (1997b). "A Constitution for Knaves Crowds out Civic Virtues." *Economic Journal* 107 (443): 1043-1053.

Frey, Bruno S. and Eichenberger, Reiner. (1995). "Competition among Jurisdictions: The Idea of FOCJ." In *Competition Among Institutions,* Gerken, Luder, ed., pp. 209-229. London: Mac Millan.

Frey, Bruno S. and Eichenberger, Reiner. (1996). "FOCJ: Competitive Governments for Europe." *International Review of Law and Economics* 16 (3): 315-327.

Frey, Bruno S. and Eichenberger, Reiner. (1997a). "FOCJ: Creating a Single European Market for Governments." In *Constitutional Law and the European Union,* Schmidtchen, Dieter and Cooter, Roger, eds. Aldershot: Edward Elgar.

Frey, Bruno S. and Eichenberger, Reiner. (1997b). "Combining Federalism and Democracy in Developing Countries." *IB Review* 1 (1): 3-9.

Garret, Geoffrey. (1992). "International Cooperation and Institutional Choice: The EC's Internal Market." *International Organization* 46 (2): 533-560.

Gwartney, James D. and Wagner, Richard E. (eds.) (1988). *Public Choice and Constitutional Economics.* London: JAI Press.

Hochman, Oded, Pines, David, and Thisse, Jacques-François. (1995). "On the Optimal Structure of Local Governments." *American Economic Review* 85 (5): 1224-1240.

Holcombe, Randall. (1991). "Constitutions as Constraints: A Case-study of Three American Constitutions." *Constitutional Political Economy* 2 (3): 303-328.

Hotelling, Harold. (1929). "Stability in Competition." *Economic Journal* 39 (1): 41-57.

Inman, Robert P. (1999). "On designing Intergovernmental Transfers with an Application in the New South Africa." In *Environmental and Public Economics Essays in Honor of Wallace E. Oates,* Panagariya, Arvind, Portney, Paul R., and Schwab, Robert M., eds., pp. 222-252. Cheltenham: Edward Elgar.

Inman, Robert P. and Rubinfeld, Daniel L. (1996). "Designing Tax Policy in Federalist Economies," *Journal of Public Economics* 60 (3): 307-334.

Inman, Robert P. and Rubinfeld, Daniel L. (1997a). "Rethinking Federalism." *Journal of Economic Perspectives* 11 (4): 43-64,

Inman, Robert P. and Rubinfeld, Daniel L. (1997b). "The Political Economy of Federalism." In *Perspectives on Public Choice,* Mueller, Dennis C., ed., pp. 73-105. Cambridge: Cambridge University Press.

Josselin, Jean-Michel and Marciano, Alain. (1995). "Constitutionalism and Common Knowledge: Assessment and Application to a Future European Constitution." *Public Choice* 85 (1-2): 173-188.

Josselin, Jean-Michel and Marciano, Alain. (1997). "The Paradox of Leviathan: How to Develop and Contain the Future European State." *European Journal of Law and Economics* 4 (1): 5-21.

Josselin, Jean-Michel and Marciano, Alain. (1999a). "Administrative Law and Economics." In *The Elgar Companion to Law and Economics,* Backhaus, JürgenG., ed., pp. 115-120. Aldershot: Edward Elgar.

Josselin, Jean-Michel and Marciano, Alain. (1999b). "General Norms and Customs." In *The Elgar Companion to Law and Economics,* Backhaus, Jürgen G., ed., pp. 233-239. Aldershot: Edward Elgar.

Josselin, Jean-Michel and Marciano, Alain. (1999c). "Unitary States and Peripheral Regions: A Model of Heterogeneous Spatial Clubs." *International Review of Law and Economics* 19: 501-511.

Josselin, Jean-Michel and Marciano, Alain. (2000). "Displacing Your Principal. Two Historical Case Studies of Some Interest for the Constitutional Future of Europe." *European Journal of Law and Economics* 10 (3): 213-233.

Josselin, Jean-Michel and Marciano, Alain. (2001). "Public Decisions in the Scottish Enlightenment Tradition." *Journal of Economic Studies* 28 (1): 5-13.

Kalt, Joseph P. and Zupan, Mark A. (1984). "Capture and Ideology in the Economic Theory of Politics." *American Economic Review* 74 (3): 279-300.

Kau, James B. and Rubin, Paul H. (1979). "Self-interest, Ideology and Logrolling in Congressional Voting." *Journal of Law and Economics* 22 (2): 365-384.

Kincaid, John. (1998). "The Devolution Tortoise and the Centralization Hare." *Federal Reserve Bank of Boston—New England Economic Review* (May/June): 13-40.

Kirchner, Christian. (1997). "Competencies Catalogues and the Principle of Subsidiarity in a European Constitution." *Constitutional Political Economy* 8 (1): 71-87.

Koslowsky, Rey. (1999). "A Constructivist Approach to Understanding the European Union as a Federal Polity." *Journal of European Public Policy* 6 (4): 561-578.

Kornai, Janos. (1976). "The Soft Budget Constraint." *Kyklos* 39 (1): 3-30.

Lenaerts, Koen. (1990). "Constitutionalism and the Many Faces of Federalism." *American Journal of Comparative law* 38 (2): 205-263.

Mancini, Federico. (1991). "The Making of a Constitution for Europe." In *The New European Community,* Kehoane, Robert and Hoffmann, Stanley, eds., pp. 177-194. Boulder: Westview press.

Marciano, Alain and Josselin, Jean-Michel. (eds.) (2002). *The Economics of Harmonizing European Law*. Cheltenham: Edward Elgar.

McConnel, Michael. (1987). "Federalism: Evaluating the Founders' design." *University of Chicago Law Review* 54:1484-1512.

Merville, Larry J. and Osborne, Dale K. (1990). "Constitutional Democracy and the Theory of Agency." *Constitutional Political Economy* 1 (3): 21-47.

Musgrave, Richard A. (1997). "Devolution, Grants, and Fiscal Competition." *Journal of Economic Perspectives* 11 (4): 65-72.

North, Douglas C. (1990). *Institutions, Institutional Changes and Economic Performance*. Cambridge, New York: Cambridge University Press.

Nozick, Robert. (1974). *Anarchy, State and Utopia*. New-York: Basic Books.

Oates, Wallace E. (1972). *Fiscal Federalism*. New York: Harcourt, Brace, Jovanovitch.

Oates, Wallace E. (ed.). (1998). *The Economics of Fiscal Federalism and Local Finance*. Aldershot: Edward Elgar.

Oates, Wallace E. (1999). "An Essay on Fiscal Federalism." *Journal of Economic Literature* 37 (3): 1120-1149.

Ostrom, Vincent. (1976). "The American Experiment in Constitutional Choice." *Public Choice* 27 (3): 1-12.

Pauly, Mark. (1973). "Income Redistribution as a Local Public Good." *Journal of Public Economics 2* (1): 33-58.

Peltzman, Samuel. (1980). "The Growth of Government." *Journal of Law and Economics* 23 (2): 209-287.

Peltzman, Samuel. (1984). "Constituent Interests and Congressional Voting." *Journal of Law and Economics* 27 (1): 181-210.

Qian, Yingyi and Weingast, Barry R. (1997). "Federalism as a Commitment to Preserving Market Incentives." *Journal of Economic Perspectives* 11 (4): 83-92.

Rose-Ackerman, Susan. (1981). "Does Federalism Matter? Political Choice in a Federal Republic." *Journal of Political Economy* 89 (1): 152-165.

Salmon, Pierre. (1987). "Decentralisation as an Incentive Scheme." *Oxford Review of Economic Policy* 3 (2): 24-43.

Samuelson, Paul A. (1954). "The Theory of Pure Public Expenditures." *Review of Economics and Statistics* 36 (6): 387-389.

Sandler, Todd and Tschirhart, John. (1997), "Club Theory: Thirty Years Later." *Public Choice* 93 (3-4): 335-355.

Schilling, Theodor. (1995). "Subsidiarity as a Rule and a Principle, or Taking Subsidiarity Seriously." www.law.harvard.edu/prograras/JeanMonnet/papers/95/9510ind.html.

Schilling, Theodor. (1996). "The Autonomy of the Community Legal Order: An Analysis of Possible Foundations." *Harvard International Law Journal* 37.

Sobel, Russel S. (1999). "In Defense of the Articles of the Confederation and the Contribution Mechanism as a Means of Government Finance: A General Comment on the Literature." *Public Choice* 99: 347-356.

Starrett, David. (1988). *Foundations of Public Economics*. Cambridge: Cambridge University Press.

Stein, Eric. (1981). "Lawyers, Judges and the Making of a Transnational Constitution." *American Journal of International Law* 75: 1-27.

Stone Sweet, Alec. and Caporaso, James A. (1998). "La Cour de Justice et l'intégration Européenne." *Revue Française de Science Politique* 48: 195-244.

Tannenwald, Robert. (1998). "Devolution: The New Federalism—An Overview." *Federal Reserve Bank of Boston—New England Economic Review* (May/June): 1-12.

Tiebout, Charles M. (1956). "A Pure Theory of Public Expenditures." *Journal of Political Economy* 64 (5): 416-424.

Tullock, Gordon. (1969). "Federalism: Problems of Scale." *Public Choice* 6: 19-29.

Vibert, Frank. (1995). *Europe: A Constitution for the Millenium.* Aldershot: Dartmouth.

Wagner, Richard E. (1988). "Agency, Economic Calculation and Constitutional Construction." In *The Political Economy of Rent-seeking,* Rowley, Charles K., Tollison, Robert D., and Tullock, Gordon., eds., pp. 423-445. Boston: Kluwer.

Wagner, Richard E. (1993). *Parchment, Guns and the Maintenance of Constitutional Contract. Shafestbury Papers 3.* Aldershot: Edward Elgar.

Wallis, John J. (1999). "Early American Federalism and Economic Development, 1790–1840." In *Environmental and Public Economics. Essays in Honor of Wallace E. Oates,* Panagariya, Arvind, Portney, Paul R., and Schwab, Robert M., eds., pp. 283-309. Cheltenham: Edward Elgar.

Weiler, Joseph H. (1991). "The Transformation of Europe." *Yale Law Journal* 100: 2410-2431.

Weiler, Joseph H. and Ulrich R. Halpern. (1996). "The Autonomy of the Community Legal Order. Through the Looking Glass." *Working Paper, Harvard Law School.*

Weingast, Barry R. (1995). "The Economic Role of Political Institutions: Market-preserving Federalism and Economic Growth." *Journal of Law, Economics and Organization* 11 (1): 1-31.

Weingast, Barry R. (1997). "Democratic Stability as a Self-enforcing Equilibrium." In *Understanding Democracy. Economic and Political Perspectives,* Breton, Albert, Galeotti, Pietro, Salmon, Pierre, and Wintrobe, Ronald, eds., pp. 11-46. Cambridge: Cambridge University Press.

Wittman, Donald. (1991). "Nations and States: Mergers and Acquisitions; Dissolution and Divorce," *American Economic Review* 81 (2): 126-129.

Zodrow, George and Mieszkowski, Peter. (1986). "Pigou, Tiebout, Property Taxation, and the Under-provision of Local Public Goods*." Journal of Public Economics* 19 (3): 296-315.

Chapter 18

FISCAL SOCIOLOGY: WHAT FOR?

Jürgen G. Backhaus
University of Erfurt
juergen.backhaus@uni-erfurt.de

Abstract   In discussing the question: fiscal sociology: what for, I shall first give a short sketch of the history of thought of the field. Secondly, main issues are identified. First in discussing the concept of the tax state, we emphasize issues in constitutional public finance. Secondly, one of the fields in which fiscal sociology has been most important is the issue of taxation, and notably income taxation. Thirdly, in citing applications and issues, an entire alphabet of fiscal sociological issues is being identified. The third paragraph deals with the future of the field in both instruction and research.

Keywords: Rudolf Goldscheid, Vilfredo Pareto, Joseph Alois Schumpeter

JEL classification: B1, B2, H0, Z0

In discussing the question: fiscal sociology: what for, I shall first give a short sketch of the history of thought of the field. Secondly, main issues will be identified. First, in discussing the concept of the tax state, we emphasize issues in constitutional public finance. Secondly, one of the fields in which fiscal sociology has been most important is the issue of taxation, and notably income taxation. Thirdly, in citing applications and issues, an entire alphabet of fiscal sociological issues is being identified. The third paragraph deals with the future of the field in both instruction and research.

1.   A SHORT HISTORY OF FISCAL SOCIOLOGY

Fiscal sociology as a separate field makes sense only once economics and sociology have parted ways, leaving a void in between. Originally, when economics started as a separate field of instruction and research on the European continent, there was no need for a specialty in fiscal sociology. In repeating a

perhaps familiar story, economics in an institutional sense started on the European continent when King Frederick William of Prussia, being dissatisfied with the state of the instruction of his future civil servants, founded two chairs in cameral sciences at the universities of Halle (Saale) and Frankfurt (Oder) in 1723. Cameral sciences meant, in essence, public finance. Hence, on the European continent, economics as a separate discipline started with public finance as the core subject area, since the purpose of economic thinking and teaching was to develop institutions so as to foster the wealth and happiness *(Wohlstand und Glückseligkeit)* of the states and thereby the people living in the house of the state. Hence, the wealth of a nation was seen as the result of prudent state economic policy. It would take another 99 years before at Oxford the first chair in political economy was founded in Britain. Here, the question of the nature and causes of the wealth of nations (1786) had been posed rather differently by Adam Smith. He saw the causes as lying in the division of labour as a precondition for extensive trade; the state would be reduced to taxation and upholding the public order, but was not seen as an active participator in economic activity. These two views of economics prevail to this very day, and they are often seen in different positions taken with respect to economic policy undertaken by the European Union.

While on the European continent, ever-larger numbers of students had been trained in economics, the instruction in the field was late to come in Britain and the United States. In Britain, the introduction of the tripos by Alfred Marshall brought a very small number of students indeed. In the United States, it was notably the emerging activity of Washington-based agencies in matters of social policy which required a larger number of well-trained economists and statisticians, i.e., a development of the 20th century. When from 1723 onwards, economics was taught on the European continent, it was done in the context of cameral and policy sciences and the law. Cameral sciences were essentially economics and public finance, all this embedded in a broad social policy approach with the state and its institutions as the main source of initiative. Next to this, the student would be instructed in public administration *(Policeywissenschaften)* and law, with a strong element of technology-related issues covered as well. These might have to do with mining, agriculture, forestry or manufacturing. This entire field of related areas was called *Staatswissenschaften,* and many universities therefore had their own faculties of *Staatswissenschaften,* binding these fields of what today is law, economics, public administration, political science, sociology and contemporary history and policy together. This organization of instruction and research explains why the literature in economics and public finance all through the 19th century has a strong multi-disciplinary and interdisciplinary character, certainly as seen from today's point of view. Even in the 20th century, this is still true for the work of such authors as Werner Sombart, Joseph Alois Schumpeter or Max

Weber who, while pioneering in sociology, did not consider themselves primarily sociologists. They all embody the fusion of economics and sociology: in Sombart's case, with economic history in addition, in Schumpeter's case with his strong push for economic analysis and econometrics and in Weber's case with the additional integration of law, his original field of expertise. However, with the pioneering work of Emile Durkheim, Vilfredo Pareto, Georg Simmel and Ferdinand Tönnies, the separate field of sociology had emerged.[1]

These separate developments in sociology had the effect that the two disciplines, economics on the one hand and sociology on the other, parted ways. With the rapid development of the social sciences in the 20th century, economics including public finance, business economics, public administration, sociology, political science and policy science all came to exist side-by-side with separate research programs, and the links between all of them on the one hand and law on the other became more and more severed.[2]

The differentiation of separate disciplines within what formerly had been *Staatswissenschaften* left lacunae or gaps in between. While methodological eclecticism had reigned within the *Staatswissenschaften* with very few exceptions,[3] now the different disciplines within the social sciences also differentiated themselves by the methods used and, from a methodological point of view, became more homogeneous internally, but heterogeneous in distinguishing one from the other. The logical consequence was a formation of clusters linking subjects with the appropriate methods. While some subjects remain amenable to an analysis by different methods, others became neglected. As we compare, for instance, the tables of contents of two Chicago-based journals, the *Journal of Political Economy* and the *American Journal of Sociology,* we notice that many topics recur in both publications, such as issues relating to the family, issues relating to behaviour in the workplace or issues relating to the organization of firms and enterprises. Remarkably different are, of course, the questions asked and the methods of analysis used, although both journals emphasize empirical work. On the other hand, the large comprehensive studies leading to ambitious projects of legislation, which characterized the activities of leading social scientists towards the end of the 19th and the beginning of the 20th century by the end of 20th century, have become de-emphasized. The reason is easy to see. Such large legislative enterprises as the launching of a welfare state, the codification of the civil code, a commercial code or the formation of a system of public law or variously the construction of a workable comprehensive national system of health care all require carefully combined eclectic methods synchronized with respect to the questions asked, the empirical data generated and used, and the policy advice generated so as to be fit for legislation and implementation.

That the differentiation in the social sciences had led to a gap which needed to be filled with a fresh approach became clear towards the end of World War

I, when the system of war finance had not only destroyed the state institutions of the allied powers of Germany and Austria to the core; next to the immense human and material losses, the defeated states could not continue their operation after the war in the way they had done before. This was first pointed out by Rudolf Goldscheid, who is also the father of the term fiscal sociology *(Finanzsoziologie)* and therefore one of the first classical authors in the field. His contribution has to be seen next to that of Joseph Schumpeter, since Schumpeter in his classic piece "The Crisis of the Tax State" responded to Goldscheid's analysis and thereby contributed the second classical piece to the field. To this debate, we should now turn.

2.   MAIN ISSUES

2.1.   The Debate on the Crisis of the Tax State

The roots of contractarian or constitutional reasoning are to be found in the tradition of liberalism. The doctrines of liberalism, insofar quite different from the present Leviathan approaches centred around two related sets of civil rights for which constitutional guarantees were sought as a means of protection against the discretion of the king or sovereign: this is because as far as the absolute ruler is concerned, property and the domain of individual liberties are substitutes; in principle, for each infringement on individual liberties imposed by a ruler to further some of his interests, there is a functional tax equivalent with which these interests can be equally served, and *vice versa.* This broader view is also taken into account by Goldscheid, who accordingly advanced an evolutionary theory of the state where, in the beginning, the state as personified by the prince could seek either revenues or services in kind.

Goldscheid's theory systematically relies on this dualism, upon which a second dichotomy is constructed. There are two classes of citizens in the population, the owners of labour and the owners of capital. While the state is able to tax the former and, beyond the point of optimal tax extraction, demand services in kind, capital, which is more flexible and powerful in his model, is only borrowed, and the state incurs the public debt. Where the first group actually contributes to the state's expenses through taxes or (mostly military) services, the second receives a claim in return, a claim which has to be satisfied later out of the general tax revenue or by services in kind.

These services, in particular the draft, play a crucial role in Goldscheid's politico-economic analysis, which is also designed to explain the extraordinary length of duration of World War I. Both Goldscheid[4] and Schumpeter[5] agreed that some state activities, such as the war, could never have been carried out had the enormous cost immediately and visibly been shifted to an identifiable public through ex-propriative taxes. While Schumpeter, however, argued more technically in terms of the maximum exhaustibility of the tax base, Goldscheid

put forward his interest group perspective, where the creditors to the public (capital owners) had no interest to end the war, never expecting to be required to foot the bill, rather receiving reliable promises to be repaid.

In either case, the war debt contributed to **fiscal illusion** in that it covered up the destruction of real resources and property and so helped carry out policies which, had their true costs been obvious to the citizens, would never have been accepted. When honouring the war debt, in Goldscheid's model, the labour class would end up with the entire bill as the creditors to the public demanded their interest and repayment out of the national dividend.

It is here where Goldscheid's peculiar approach to the notion of human capital *(Menschenökonomie)* becomes relevant. Whereas the contributions of the labour class to the expenses of the state represent real goods and services, either taxes which represent part of the national product, services in kind or other infringements on individual liberty, the capital class contributes only credits, which are to be repaid. Thus, even when real capital, human and material, is used up or destroyed, only the owners of material capital continue to present claims. These claims constitute political leverage, which, still according to Goldscheid, the creditors use to have the state governed in their own interest.

Whereas this model was obviously constructed under the impression of politico-economic interaction during the First World War in Austria and Germany, Goldscheid's proposal for fiscal reform was based more broadly on his unorthodox interpretation of the fiscal history of the state, from the middle Ages to his days. This story begins with a strong state, independent and relatively rich, relying on large property holdings. Only after the Thirty Years' War, however, the rising demands of the budget exercised by war finance and the desolation of the country give rise to a new approach to economic policy: cameralism, where the state, gradually transforming itself into the tax state, follows policies of economic development in order to strengthen the tax base, notwithstanding engagement in traditional and new forms of public entrepreneurship; both types of policies aiming at long run revenue maximization. This dual policy is constrained by the two relationships governing fiscal technology as discussed above: first, the Cameralist relationship as the interdependence between public spending and the productivity of the tax base; and second the experience of rising marginal costs of tax extraction, which leads to the definition of a point of optimal extraction (Laffer curve).

The issue of defining a tax constitution has re-emerged on the political agenda with the European Union beginning to take shape and with elements of an emerging European constitution becoming visible. These tasks, which we face today, cannot be solved in a technical manner based on simple models. The fiscal constitution of the European Union will have to meet on the one hand the classical criteria developed in political doctrine, it has, on the other hand, to be in line with the requirements of a modern global economy, and

thirdly it has to, not only accommodate member states with very different constitutions, economic and political systems, including different systems of taxation and political decision-making, but also different histories, cultures and socio-political visions. In order to meet this challenge, the different disciplines of public finance, law, public administration, political science and sociology will have to join forces as they had in the tradition of *Staatswissenschaften.* In order for such a joint venture to be successful, integrative paradigms such as those of the tax state may very well be useful in the future as they have been in the past.

2.2.   Income Taxation

At a conference of the international Schumpeter Society in Kyoto (Japan),[6] Richard A. Musgrave, a student of Schumpeter's surprised his audience by insisting that Schumpeter's contribution to public finance had been minimal. He only emphasized, and actually overemphasized the distortions caused by income taxation. In fact, Schumpeter's name rarely appears in current textbooks on public finance.[7] Indeed, Schumpeter's extensive contributions to public finance[8] tend to be collected in his sociological or political writings.[9]

In this sense, Musgrave is perhaps correct that most of these writings do not fit within his three-winged cathedral of public finance dealing with allocation, distribution, and stabilization. Schumpeter stood firmly in the continental European tradition of public finance, from Puviani with his emphasis on fiscal illusions to Wicksell[10] with his emphasis on taxation and political decision-making and Da Empoli's[11] emphasis on the multiple economic and social distortions caused by all manner of taxation. In this sense, Schumpeter never severed the ties between public finance and the neighbouring disciplines and can therefore be properly claimed as one of the fathers of fiscal sociology. Indeed, Musgrave's hint may serve to offer a simple explanation of what fiscal sociology is about.

2.3.   Applications and Issues

In the simplest of cases, consider a product (bagels) and an excise tax levy on this product.

The excise tax increases the demand price, it decreases the supply price and the equilibrium quantity decreases from Q_0 to Q_t. The excess burden is the shaded triangle *bad,* which is equal to the excess burden of the marginal willingness to pay (for bagels) over the marginal costs (of bagels) for those bagels not produced and consumed because of the tax. Here, the textbook treatment ends. Yet the questions remain: What did the people eat who did not eat those bagels that have not been produced and consumed? What did the baker do while not producing the bagels? What, in fact, did consumers and bakers hatch
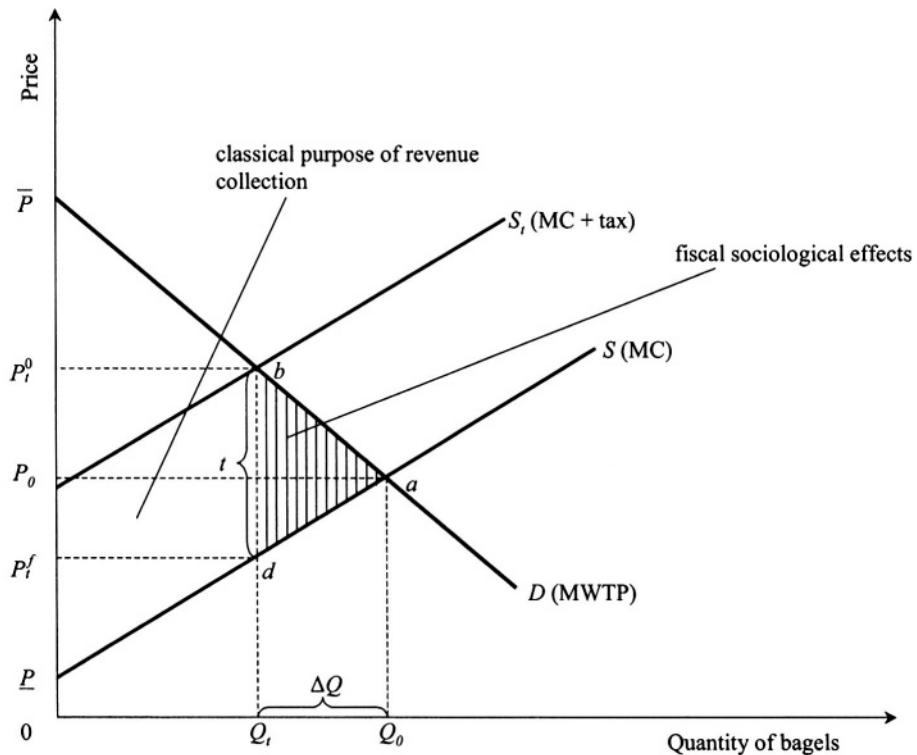
FIGURE 1.

as new ideas so as to be able to avoid the tax? Shall we see new products on the markets to which the tax does not apply? Can we discern new forms of distribution to which the tax would not apply? Under what circumstances can those forms of tax avoidance arise? Where do they thrive and what further consequences do they entail? All these questions can not be asked within the traditional concept of tax burden analysis, but all these questions are indeed those we are interested in when thinking about levying, increasing or decreasing a particular tax.

This standard example of excess burden analysis in the partial and static form just presented falls far short of what Schumpeter tried to attempt in the seventh chapter of his "Theory of Economic Development," which is not contained in the English translation nor in the subsequent editions of the German text.[12] Building on the distinction between static and dynamic analysis, partial and total (in the sense of comprehensive) modelling and introducing entrepreneurial change (defined as new compositions) which affects technology (i.e., the production function) and through this all the other factors involved, Schumpeter tries to describe (verbally) the enormous complexity that would face the

economic modeller in the Lausanne tradition. For good reason, he did not try to write such models, but rather tried to use his considerable rhetorical talent to argue a case for a complex analysis, where, indeed, the argument is always on two levels, a simple and a complex one within the same language structure. Even today, with computer-supported modelling strategies we have not been able to come up with credible alternatives to Schumpeter's flowering language on various levels, akin to the suspended gardens of Semiramis. Both instruction and research in fiscal sociology will have to take heed from Schumpeter's attempt, both where he succeeded and where he obviously failed. Before addressing such strategies of instruction and research, however, let us turn to a subjective list of issues interesting in the field of fiscal sociology, and let us try to exhaust the Western alphabet.

2.4.   An Alphabet of Fiscal Sociology

The existence of *alternatives* or different opportunities to choose from is at the heart of the possibility of doing economics as a science.[13] Only if there are alternatives to choose from can there be any reaction to an action of the government facing a taxpayer. From the point of view of public policy, the difficulty consists in a central authority's inability to know the full extent of the choices from which a citizen can choose who is intent on minimizing the impact of a governmental action on his own income or wealth. This knowledge about the alternative ways of avoiding harm, such as the burden of taxation (including both the tax itself and the excess burden) is available only in the decentralized form of knowledge and households, firms and other economic entities. It cannot be systematically collected by any central authority.[14] It is for this reason that knowledge about the excess burden will be systematically underestimated by any conceivable attempt at measurement.

The *burden* of taxation can be properly identified as the focus of research in fiscal sociology. The difficulty with this subject consists in its elusiveness. As citizens try to minimize the impact of the burden of taxation (and regulation), they invent ever-new forms of legal avoidance. Once this interaction is properly understood, the anticipation of such avoidance behaviour can be the cornerstone of public policy itself. By explicitly including loopholes into tax codes and regulations, citizens or firms can be made into agents of public policy.

For instance by granting tax *credits* for certain favoured investments such as ecologically preferred technologies, a shift in technological development into a desired direction may be accomplished.[15]

Depreciations, in particular accelerated depreciations can be used to a similar effect. Reginald Hansen[16] has documented extensively how (through the exemptions introduced into the income tax code under Section 7) German authorities were able to rebuild the stock of German housing after World War

II, rebuild a commercial fleet and essentially rebuild Eastern Germany after reunification. Instead of enacting all these directly, use is made of the initiative of private individuals and their desire to reduce the burden of taxation. This approach has two redeeming consequences. On the one hand, the financial resources spent on, for instance, rebuilding the historical core of a destroyed city according to municipal zoning guidelines do not have to be raised through taxation, hence there is no excess burden. Secondly, opportunities for realizing these preferred investments will be found by private entrepreneurs who can use knowledge that would not be available to a government authority either in the quality or to the extent as it is available privately. Although the method looks roundabout, it is more direct than a tax-financed government programme, which would be its alternative.

The *elasticity* of a particular activity, such as the demand for bagels in the illustration given above, determines the extent or size of the excess burden and, by implication, the realm of the fiscal sociological dimensions of a particular policy measure. Only if a particular activity, such as the demand for heating oil, is completely inelastic, will there be no excess burden, no opportunity to reduce or avoid the burden of the tax. Such cases of inelastic demand or supply functions are very rare indeed. And this implies, again, that the opportunities for avoiding the burden of taxation (or regulation) are manifold and hard to anticipate by a government agent.

Fiscal illusion is the extent to which the burden of a tax (or regulation) is underestimated by a citizenry or, conversely, the benefits of a particular government programme are overestimated. It is in the interest of a government to try to achieve fiscal illusion, and this is true both for democratic governments that are in need to win re-elections, as for non-democratic regimes that likewise need to maintain their grip on power.[17] According to the canons of taxation, fiscal illusion should, of course, be avoided. However, given the incentives on the part of taxing and regulatory authorities, this rule is certainly not self-enforcing. If we do not make a special effort at discovering fiscal illusion, it will remain undetected and uncorrected, with the consequence of inefficiency in the public sector: too many projects that do not fit the demand of the citizenry, and too high a burden of taxation with the concomitant high excess burden, i.e., welfare loss. In this sense, research into fiscal illusion can be welfare enhancing in uncovering government caused inefficiencies.

Government bonds can be equivalent to taxation when the strict conditions of the Ricardian equivalence theorem hold. In that case, the anticipated future burden of taxation needed to service and redeem the bond will be factored into the decisions of current taxpayers who rationally perceive their wealth as being reduced by the future burden of taxation. However, these strict conditions only rarely hold.[18] A conceivably extreme case may illustrate the point. Starting in the 80s of the last century, an extensive net of charitable foundations had

been set up in Germany, many serving to finance scholarly work, even entire universities such as the University of Frankfurt (Main). These foundations had to hold their wealth in government bonds. As a consequence, after World War I these foundations had lost their wealth. Imperial government bonds could no longer be redeemed, including those issued during the war. The hyperinflation resorted to after the *coup d'état* completed the financial destruction of these foundations. In this case, the consequences, i.e., the harm done in terms of opportunities foregone for this charitable and scientific work that would otherwise have taken place through these foundations is extremely difficult to be accurately established. It can probably only be described, it may not be possible even to attempt a measurement. We thus realize that the limiting conditions of the Ricardian equivalence theorem point to a vast area of research for fiscal sociologists into the burden of government bond issue.[19]

Health effects of taxation can be manifold and surprising. In an effort at harmonizing European excise taxes, Spain departed from its traditional tobacco monopoly which distributed tobacco products through licensed shops and introduced an excise tax along with a liberalization of tobacco sales. In particular the international producers of low-tar brands had pushed for this reform, which was also in line with European harmonization efforts. The surprising result, however, was the disappearance of the traditional local tobacco shops *(estancas)* and the appearance of contraband tobacco sales in bars and restaurants. Obviously, this resulted in heavy losses in public revenue; but it also resulted in an increase in the consumption of high-tar tobacco products, which have more adverse health effects than low-tar tobacco products. Without understanding the precise consequences, an established system of taxation (through monopolization) and containment of consumption had been displaced, with the resulting loss in public revenue and an increase in the consumption of the product for which a temperance policy had been envisaged.

Income in kind programmes can likewise have surprising consequences. The socialist Allende administration (1970–1973) in Chile had promised to deliver a litre of milk to every child. It started introducing price controls on dairy products, which, along with the heavy inflation that set in shortly after the administration had taken over, first led to a glut in meat and consequently to a disappearance of dairy products. The administration then bought milk powder in international markets, and distributed bags of milk powder to parents of children under the programme. Due to the enormous shortage of dairy products, these milk powder bags had a substantial resale value and helped the families suffering under the inflationary devaluation of their wage income to support this income through sales in the black market. The outcome, of course, was that the milk did not end up feeding the children targeted by the programme. Such a scenario is, perhaps, extreme, but it follows a strict eco-

nomic logic and is clearly predictable. Whenever aid is granted in kind, one has to look at both the income and substitution effects.[20]

Jointness of investment and returns is an important principle in ensuring that a particular economic activity is seen in its entirety and taxed accordingly. This principle is often violated when income is taxed according to the source principle without consideration of the expenses it affords. Schmoller had insisted that income be measured as that amount that can be consumed without reducing wealth. Hence, if a household consumes € 100,000 and its wealth had been € 10 million in period 1 and € 10 million in period 2, this consumption is the income that has to be taxed. However, if you disjoin consumption and expenditure behaviour from income and revenue behaviour, interesting effects can occur. The case has been documented of the owner of a landed estate who lived lavishly but never realized any income whatsoever. Hence, he never paid any income tax. In a suggestive way, this can work as follows. Imagine you own a resource which is valued at € 10 million. It should increase in value at some rate, and you can therefore increase your borrowing against the asset at the same rate, discounting for interest. If the asset, for instance, is attractive land that benefits from development of bordering tracts of land, the (speculative) increase in value will be much more than the normal market return on bonds, and you can borrow against this asset accordingly. This may again support this lavish lifestyle, while no income is generated, rather further and further debt, when the asset that increases in value serves as collateral. If you insist on defining income in terms of its source instead of what it allows in supporting your lifestyle, that is, if you insist in disjoining assets and returns, you will end up with the paradoxical result that the lavishly maintained lifestyle nevertheless is not subject to income taxation. Only by studying the reality of taxpayer response to particular rules of taxation can we establish a realistic picture of the impact of taxation on the behaviour of people and the economic effects of these dual causes.

Kiddy tax is a phenomenon which penalizes households having children under the tax code. It can take many forms.[21] To start with, a woman having a child necessarily has to leave the workplace for a while, and it depends on the cultural environment how long this 'while' will actually take. It may be three months, it may be three years, and there are cultures where the process takes longer. This means, however, that the mother may lose not only her proficiency in her job, but also the regular salary increments going along with normal performance. She will thus start at where she stopped, and carry the entire loss forward over the expected duration of her labour experience. From the point of view of the employer, this cannot be otherwise, since the wage has to reflect marginal productivity. From the point of view of taxation, however, it is quite possible to compensate for this necessary effect of having children while being employed. This compensation will have to be the higher, the more

the mother earns—an awkward result when equity in taxation is emphasized in terms of equalizing incomes and wealth. In addition, raising children brings about lots of additional expenses, which, if they are not deductible, in fact weigh in as a kiddy tax. Some recent programmes allow mothers to establish home offices.[22] Instead of granting subsidies to some mothers for such home offices, it would be more effective to grant tax deductions for home offices actually established. That would allow for mothers who see an opportunity to combine the work with childhood at home develop entrepreneurship and establish the proper working environment. It is difficult for a state agency to identify such able home working entrepreneurial mothers and tailoring a grant to the needs that actually exist.

Liability for false returns. No market can work without the institution of liability in contract and wrongdoing (violating property rights). Yet in taxation, very little is there in the way of liability for false returns. Tax returns may be used for all manner of ends, not only assessing the tax code, but also different social policy objectives. After all, tax returns are a goldmine of information, and this information, if it is correct, can be effectively used in order to tailor government programmes to citizens' needs. However, if rules of taxation are not self-enforcing, the information will be totally misleading and it will not be useful for any ulterior government purpose. Since liability for false returns is difficult to be made actionable (without reversing burdens of proofs), the self-enforcing mechanism practiced in some Swiss cantons is instructive. Here, the method of self-declaration of income is being practiced, but the returns are made public and banks are prudent in using them for their crediting decisions, lest they be criticized in public themselves. This means that a local entrepreneur cheats on his taxes only at the expense of losing his credit line. Banks will deal only with local customers, and the system is closed.

Monitoring compliance is not only important with respect to honesty in taxation. It is particularly difficult for the tax authority when a tax is used as a policy instrument so as to achieve a particular purpose, such as the use of an ecologically preferred technology, production process or resource. The information required to monitor compliance may become very complex in such cases, and it is important to create a system of self-enforcement. Often, the necessary expertise cannot be assumed to lie with a taxing authority, certification may be required from a different agency. In such cases, the extensive use of tax credits reverses the burden of providing the information from the taxing authority to the taxpayer, who thereby has to prove compliance. This reversal of the burden of proof expedites the tax administration and offers a flexible instrument to achieve many and diverse policy objectives.[23]

Nominal gains or losses (or even fictitious or forfeited incomes) are sometimes used for assessing taxes. In principle, only real entities will affect economic behaviour. However, when used for purposes of assessing taxes, fic-

titious values do indeed become real ones. In this sense, they can affect behaviour and by consequence have an influence on real wealth and welfare. Consider the case of the fictitious value of using one's own home. Under the Dutch tax code, for instance, the forfeited benefit of living in one's own home is calculated not on the basis of rents for equivalent leases, but on the basis of the value of the real estate. This figure is then added to the taxable income, adding to the progressive burden of the income tax. In principle, from a social policy point of view, the objective is to have as much homeownership as possible. This has to do with provisions for old age or other vicissitudes of life, with the stability of neighbourhoods, with health considerations and with the general experience that privately owned housing stock is better and more efficiently maintained than publicly or institutionally owned one. Now consider a family with a modest income that has inherited a valuable house. The fictitious rental value of the house will be added to the income of that family, but the costs of maintaining the house, the opportunity costs of capital contained in the real estate and the opportunity costs of work upon the house are all not deducted from this forfeited income. Consequently, the deck is stacked against private homeownership, thereby frustrating attainment of the traditional social policy objective. It is unlikely that this result is based on a conscious decision of parliament.

Off-shore activities are a clear indication of a weak tax administration as they document opportunities of taxation forgone. The remedy lies in creating a climate in which the business can thrive on-shore and simultaneously serve as a basis for taxation.

Public works intended to overcome chronic unemployment or other forms of economic depressions are often funded from tax revenues, thereby adding to the burden of taxation and to the welfare loss of taxation. Even if the measures are effective in creating opportunities for employment, they at the same time through the method of funding chosen decrease similarly opportunities in the sector subject to taxation and dampen demand for investment and consumption. For these reasons, self-financing forms of public works are preferable such as those designed by Wilhelm Lautenbach and implemented between 1932 and 1935. In relying on the Keynesian multiplier and accelerator effects, the funds spent wisely yield a return that constitutes the original fund.[24]

Quality of service can be tied to taxation through the benefit principle. Wicksell suggested the benefit principle, which has been a core element of the cannons of taxation ever since Justi and Adam Smith in a specific institutionalised form. By insisting on consensual taxation (near unanimity rule) and simultaneity of decisions on taxation and expenditure, taxes are tied to the benefits indeed received, and institutional provisions have to be taken to ensure that tax yields decline *in tandem* with benefits and *vice versa.* This can be accomplished by tying the taxes owned to the receipt and approval of specific

services, implying the expression of disapproval and opting out of the bundle of taxes and services.

The *Ricardian equivalence theorem* holding the equivalence between taxation and bond issue has to be taken as one of the typical theorems of limitation which form the basis of economic analysis. Similar theorems are those by Modigliani-Miller and Coase. In emphasizing the specific conditions under which the theorem holds, our attention is pointed to those cases where the conditions are not met. Hence, the equivalence theorem cannot be taken at face value. It serves to emphasize the distortions that occur when the future burden of the bond issue is not accurately perceived by the present generation or cannot adequately be compensated for in the present decisions of future taxpayers.[25]

Surprises are the focus of much work in fiscal sociology, as we try to detect and predict the unanticipated consequences of regulation and taxation. For instance, Laband[26] has recently shown that regulation designed to protect the habitat of endangered species in its effects is likely to endanger those very species. For instance, birds which nest in particular trees, if the habitat is being protected and such trees cannot be cut for logging purposes, logging firms will eliminate those trees from their forests before they attain the required age and will thereby eliminate the habitat intended to be protected. If an entire area is targeted for protection, logging firms will relinquish their activities in these areas, whereby forestry activity will stop and land will revert to other uses. In all these cases, the intended effect of the regulation is completely subverted, since the regulation is written against the interests of the regulated instead of tying the interests of the regulated and the beneficiary of the regulation together,

Time is an important element in determining the intended and unintended effects of taxation and regulation. The simple diagram shown above is completely static. In the long run, the effects of regulation and taxation cumulate and citizens learn to reduce the burden of taxation, thereby increasing the welfare loss, as the economy moves successively away from the production possibility frontier, and a similar effect occurs in polity and society.

Utility is distinguished by Pareto from ophelimity in order to distinguish between those effects which occur in the economy and those further repercussions which affect society in different ways. In principle, Pareto thought that ophelimity could be measured, while utility is in the eye of the beholder. Utility can be experienced in terms of oneself, other members of society as well as society's situation and development as such. About the latter very little can be done by an individual, and individuals may disagree strongly among each other. Further, an individual's ability to influence the utility he ascribes to the well-being of another is equally constrained. What can be objectively observed is how individuals try to improve their own well-being, and hence

the ophelimity can, in principle, be inferred, at least in an ordinal way. As we move beyond the economic sphere into the political and social effects of regulation and taxation, we need to work with both the notion of ophelimity and the different notions of utility as Pareto had suggested.

Value and votes are closely connected, as in the polity voting is used to determine the outcome of public policy, including the welfare effects of taxation and regulation. While in their individual behaviour, voters try to reduce the burden of taxation and regulation, this may not necessarily be the case in their political behaviour as the benefits from a particular programme may be concentrated on some, while the burden can be spread over many and in such a way that it is not worth their while to mount an effort at defeating the burdensome programme. Votes are the currency of the political process, and it takes real resources to muster the political strength as expressed in votes to launch or defeat particular programmes.

Wealth, its production, enhancement and preservation depend critically on very specific aspects of the tax system. This has already been noted in the context of the taxation of fictitious incomes. In fact, in the example chosen the tax structure affects the architectural heritage of the people. Likewise, in assessing wealth, human capital and its formation needs to be included along with all manifestations of the cultural heritage, such as libraries, musea, institutions of higher learning, architectural landscapes and the like. All these depend crucially on specific provisions in the tax code, such as education credits, deductions for charitable activities, deductions for maintaining national monuments, encouragement to endow foundations or chairs at universities and the like. We note that there are countless ramifications and interconnections between the tax structure and all aspects of society, which can be ascertained through a wide definition of wealth, as it has been chosen here so as to capture all manner of sociological effects of taxation and regulation.

X-inefficiency is the difference between the ability of an organization to perform and its actual performance. As an organization, a firm, a household, a family and the like tries to minimize the burden it has to bear through taxation and regulation, it moves away from its efficient allocation of resources and thereby creates X-inefficiency. Hence, to the extent that we can demonstrate these effects, we can point to strategies of reducing X-inefficiency that do not lie within the organization, but are at the disposal of policy makers as they decide on issues of regulation and taxation establishing the environment in which the organizations have to operate. While organizational theory has emphasized X-inefficiency in the context of intra-organizational strategies of improving performance, a fiscal sociological approach can emphasize extra-organizational strategies to reduce X-inefficiency.

The *yield* of taxation tends to be entirely overestimated, as the cost of taxation can never be completely established, since tax payers and agents under

regulation have no desire to cheaply provide information about their avoidance behaviour. From this point of view, it is advisable to use forms of revenue generation which strictly follow the benefit principle, as the *quid pro quo* between taxes and services provided tends to leave no excess burden and creates incentives to generate information about the transactions between the government and the private sector.

Zero-based budgeting has often been suggested as a strategy to make budgets manageable and contain the growth of government. However, as citizens learn to deal with taxation and regulation just as well as expenditure programmes, they are able to decrease distortions over time. When entire programmes or tax systems are questioned abruptly, the learning done by individuals and organizations is being undone and additional costs of transactions are being imposed on economy and society. What are required instead are budgetary procedures that incorporate as much decentralized information as possible. In using the insights from Hayek's piece on the use of knowledge in society, this can be a plea to organize as much governmental activity as possible in a market context so as to make an optimal use of knowledge in society by both public and private market participants.

3.   FISCAL SOCIOLOGY AS A FIELD OF INSTRUCTION AND RESEARCH

3.1.   Instruction

In emphasizing opportunities for research and instruction in fiscal sociology, a good start is Humboldt's principle[27] of a unity of research and teaching at the university level. Although this principle seems to fly in the face of the economic principle of the division of labour, a closer look shows that it actually reinforces this principle. Universities as institutions for basic research and high-level instruction have to maintain, increase and communicate that type of knowledge which is not readily canonized into mass education. It is necessarily the role of a research university to be small in the sense of having many small centres of learning where scholars and students can interact directly. In this sense, teaching and research opportunities can very well overlap.

As a field of teaching, fiscal sociology offers fascinating opportunities at all levels of instruction.

At the bachelor's level, elements of fiscal practice can be integrated into the course of study. Consider the bargaining situation between a regulator and a regulated activity, such as the process of applying for a building permit; or alternatively, consider negotiations after filing a tax return. These cases can be realistically simulated in the curriculum and students can be introduced into the breezing climate of adversary negotiations. Through such a method, the technical concepts such as the rectangles and the triangles shown above can

be imbued with life and may be readily grasped by students who favour rather non-technical disciplines otherwise and therefore might shy away from taking an economics course.

At the master level, this experience can be built upon. Now, the expertise would have to be raised to the professional level, since fiscal sociology includes a study of the unintended impact of taxation as well as regulation, it has implications for almost every activity of government that somehow affects the private sector. And with this, there is a wide field of professional activities in both government and the respective governmental counterpart where a proper understanding of these effects is of great use. Hence, the supply of interesting topics for master theses is inexhaustible.

Under current conditions of widespread underfunding of universities and the iron cartel of public universities with few private initiatives that can have a chance under such circumstances, it is difficult to encourage Ph.D. students to pursue an academic career. Fiscal sociology, however, is to such an extent an under-researched field that a wealth of research opportunities exists with tangible benefits to both governmental and private agents. And this should, if sufficient effort at convincing sponsors of the use of this work is undertaken, provide for research opportunities that are appealing, even to the most frugal budget administrator.

3.2.   Research

As we noted previously, before the dissolution of the unity of *Staatswissenschaften,* scholarly activity was able to centre on legislative projects, such as the civil code, the welfare state and the like. These concerted efforts can, with difficulty, be accomplished again in the context of focused niche subdisciplines. We have recently seen the emergence of health economics, which is such a niche subdiscipline in combining every conceivable economic method to analyse health related issues and institutions. Another such instance is the recent emergence of the subdiscipline of law and economics. Fiscal sociology can be positioned in a similar way. These niche subdisciplines which bridge otherwise separate scholarly pursuits can play a pivotal role in facilitating focused research on major legislative endeavours.[28]

Example 1. With the further integration of the European Union, a repetition of the large legislative project of the civil code will now have to be made for Europe. A framework legislation for a common civil code in the European Union, including harmonization and union access will have to be forthcoming and it will have to address issues of taxation and regulation while helping to integrate diverse legal institutions and cultures. Research into unintended side

effects of regulation and taxation will be of immense importance for facilitating such a project.

Example 2. The key concept of public finance is Wicksell's principle of just taxation, which entails a simultaneous process of deciding on taxes and expenditures. With the further integration of the European Union, a framework legislation will have to be forthcoming which satisfies this principle and also allows for popular decisions to be efficiently and effectively made by acknowledging, at the same time, the different histories, cultures and institutions of the different member states. Frey's[29] suggestion of functionally overlapping competing jurisdictions is one such attempt. Research into unintended consequences of such decision-making will have to be a necessary companion of launching such initiatives.

Example 3. With further integration of the European Union, some framework legislation for customs and excise taxes will have to be forthcoming that allows for the different member states and their subdivisions to remain sovereign and democratically constituted while at the same time allowing for harmonization and a minimization of the burden of those custom duties and excise taxes. Again, research into the side effects of legislation on excise taxes and customs duties will have to be necessary in order to help design efficient legislative proposals.

4.   CONCLUSION

In conclusion, fiscal sociology offers ample opportunities for research and instruction that can be said to be truly helpful in reducing the adverse effects of government activity and thereby enhancing the welfare and wealth of nations.

NOTES

1. The *Deutsche Gesellschaft für Soziologie* was founded in 1909 as an expression of specific research questions and paradigms not generally pursued by economists. For instance, Durkheim and later Franz Oppenheimer explored the connections between economic conditions and people's health states; Durkheim's were the pioneering studies on suicide. Pareto, having completed the Manual of Political Economy, saw a need to push beyond economics proper and constructed a generalized social system in his treatise translated as "The Mind and Society" by Livingston. Simmel presented in Schmoller's seminar a socio-philosophical study of money, exploring all those aspects of money that are not economic. See also: Backhaus and Stadermann (2000a) and (2000b). Sombart was interested in the conditions which led to the development of modern capitalism and in his historical studies had to go beyond economic causes, emphasizing social, political and legal preconditions for the development of capitalist institutions. Tönnies emphasized community and society, while Weber left us the gigantic attempt at Economy and Society.

2. The last instance of a close cooperation between economists and lawyers may have been the critique at the first draft of the German civil code mounted by Schmoller and Gierke, which had led to the second draft that was signed into law in 1896 and became effective—to this very day—from 1900 on. See also Backhaus (1999a) and 1999b).

3. One of them may have been the early attempt by Christian Wolff, who tried to put the entire system within his edifice of natural law. See Backhaus (1998a).

4. Goldscheid (1919).

5. Schumpeter (1942, 1976).

6. Shionoya and Perlman (1994).

7. The only exception seems to be Richard E. Wagner's public finance textbook.

8. He taught public finance at the University of Bonn from 1925 to 1932 and contributed fiscal analyses extensively to Wolfgang Stolper's "Der deutsche Volkswirt."

9. Schumpeter, Joseph A. (1918, 1953), *Das deutsche Finanzproblem.* Berlin: Der deutsche Volkswirt. Schumpeter, Joseph A. (1939), *Business Cycles.* New York: McGraw Hill, 1964. Schumpeter, Joseph A. (1943), *Capitalism, Socialism and Democracy.* London: Unwin. Schumpeter, Joseph A. (1954), *A History of Economic Analysis.* New York: Oxford University Press. Schumpeter, Joseph A. (1954), *Dogmenhistorische und biographische Aufsätze.* Tübingen: Mohr Siebeck. Schumpeter, Joseph A. (1961), *A Theory of Economic Development.* New York: Oxford University Press. Schumpeter, Joseph A. (1965), *Ten Great Economists: From Marx to Keynes.* New York: Oxford University Press. Schumpeter, Joseph A. (1967), *Economic Doctrine and Method.* New York: Oxford University Press. Schumpeter, Joseph A. (1970), *Das Wesen des Geldes.* Göttingen: Vandenbroeck & Ruprecht. Schumpeter, Joseph A. (1985), *Aufsätze zur Wirtschaftspolitik,* Seidl, Chr. and Stolper, W. F., eds. Tübingen: Mohr Siebeck. Schumpeter, Joseph A. (1992), *Politische Reden,* with comments by Seidl, Chr. and Stolper, W. F., eds. Tübingen: Mohr Siebeck. Schumpeter, Joseph A. (1993), *Aufsätze zur Tagespolitik,* with comments by Seidl, Chr. and Stolper, W. F., eds. Tübingen: Mohr Siebeck. Schumpeter, Joseph A. (2000), *Briefe/Letters,* Hetke, U. and Swedberg, R., eds. Tübingen: Mohr Siebeck.

10. Wicksell and Knut (1896).

11. Da Empoli (1931).

12. Both the original and the translation can be found at our website: www.uni-erfurt.de. Forthcoming in Backhaus (2001).

13. Buchanan (1979).

14. Hayek (1945).

15. Backhaus (1998b).

16. Hansen (1996).

17. Tullock (1987).

18. Backhaus, Holcombe and Zardkoohi (1987).

19. Backhaus (1993).

20. A classical paper is by Bruno S. Frey who showed that increases in development aid may increase weapons imports by the recipient countries. See Frey (1975).

21. Backhaus (1991).

22. Thüringer Ministerium für Wissenschaft, Forschung und Kunst, ThüringerLandeshaushalt, Kapitel 1524 Titelgruppe 84: Förderung von Frauen in Forschung und Lehre, 2000.

23. Backhaus (1998b).

24. Backhaus (1985).

25. On these limitational theorems see Backhaus (1986).

26. See Laband (2000).

27. Incidentally, this forceful reformer of cultural institutions closed the old (1392) University of Erfurt for not being willing to go with the time.

28. The leading journal in economics in the German language area during the time of these large legislative endeavours was notably called *Annals of Legislation, Administration and Economics (Jahrbücher für Gesetzgebung, Verwaltung und Volkswirtschaft).*

29. Frey and Eichenberger (1999).

REFERENCES

Backhaus, Jürgen G. (1985). "An Essay on Keynesianism in Germany." In *Keynes' Economics: Methodological Issues,* Lawson, T. and Pesaran, M. H., eds., for the *Cambridge Journal of Economics,* pp. 209-253. London: Croom Helm, Armonk, NY: Sharpe.

Backhaus, Jürgen, G. (1986). "An Impossible Economist." In *No Way: On the Nature of the Impossible,* Davis, Phil. J. and Park, David, eds., pp. 226-238. New York: W. R. Freeman.

Backhaus, Jürgen G., Holcombe, Randall G., and Zardkoohi, Asghar, (1987). "Public Investments and its Effects on the Burden of the Public Debt." *Southern Economic Journal* 54 (1): 145-158.

Backhaus, Jürgen G. (ed.) (1991). "The Economic Emancipation of Women and the Tax System." *Journal of Economic Studies* 18 (5/6). Also as book: *The Tax System and Female Emancipation.* Bradford: MCB University Press.

Backhaus, Jürgen G. (ed.) (1993). "The Economics of Science Policy: An Analysis of the Althoff System." *Journal of Economic Studies* 20 (3/4).

Backhaus, Jürgen G. (ed.) (1998a). *Christian Wolff and Law & Economics.* Hildesheim/Zürich/New York: Georg Olms Verlag.

Backhaus, Jürgen G. (1998b). "When Should the Ecotax Kick In? A Cost and Choice Approach." *Kyklos* 51 (4): 565-575.

Backhaus, Jürgen G. (ed.) (1999a). "The German Civil Code of 1896." *European Journal of Law and Economics* 7 (1): 94.

Backhaus, Jürgen G. (1999b). "Otto von Gierke." In *Elgar Companion to Law and Economics,* Backhaus, Jürgen G., ed., pp. 313-315. Cheltenham: Edward Elgar Publishing Ltd.

Backhaus, Jürgen G. and Stadermann, Hans-Joachim, (eds.) (2000a). *Georg Simmels Philosophie des Geldes: Einhundert Jahre danach.* Marburg: Metropolis Verlag.

Backhaus, Jürgen G. and Stadermann, Hans-Joachim (eds.) (2000b). *Georg Simmel's Philosophy of Money: A Centenary Appraisal.* Marburg: Metropolis Verlag.

Backhaus, Jürgen G. (2001). *Schumpeter (1884–1950): His German Contributions: The European Heritage.* Boston: Kluwer, forthcoming.

Buchanan, James M. (1979). *Cost and Choice: An Inquiry in Economic Theory.* Chicago: University of Chicago Press.

Da Empoli, Attilio. (1931). *Theory of Economic Equilibrium (A Study of Marginal and Ultra-marginal Phenomena).* Chicago: Christiani & Catenacci Publishing Co.

Frey, Bruno S. (1975). "Weapons, Exports and Aid to Developing Countries." *Journal of Conflict Resolution* 1 (2) (Spring).

Frey, Bruno S. and Eichenberger, Reiner. (1999). *The New Democratic Federalism for Europe: Functional, Overlapping and Competing Jurisdictions.* Cheltenham: Edward Elgar Publishing Ltd.

Goldscheid, Rudolf. (1919). *Grundfragen des Menschenschicksals: Gesammelte Aufsätze.* Leipzig and Vienna: E. P. Tal.

Goldscheid, Rudolf. (1919). *Sozialisierung der Wirtschaft oder Staatsbankrott: Ein Sanierungsprogramm.* Leipzig and Vienna: Anzengruber Verlag: Brüder Suschitzky.

Hansen, Reginald. (1996). *Die praktischen Konsequenzen des Methodenstreits.* Berlin: Duncker & Humblot.

Hayek, Friedrich von (1945). "The Use of Knowledge in Society." *American Economic Review* 35 (4): 519-530.

Laband, David N. (2000). "The Impact of Unfunded Environmental Mandates When Environmental Quality and Timber are Produced Jointly." *European Journal of Law and Economics* 10 (3): 199-216.

Schumpeter, Joseph Alois. (1918, 1953). *Die Krise des Steuerstaates.* Graz-Leipzig: Leuschner & Lubensky; reprinted as pp. 1-71 in: Schumpeter, J.A. *"Aufsätze zur Soziologie."* English translation by Stolper, W. F. and Musgrave, R. A. *International Economic Papers,* Vol. IV, London: Macmillan, 1954.

Schumpeter, Joseph Alois. (1942, 1976). *Capitalism, Socialism and Democracy,* 5th ed. London: Unwin University Books.

Shionoya, Yuichi and Perlman, Mark. (eds.) (1994). *Schumpeter in the History of Ideas.* Ann Arbor: University of Michigan Press.

Tullock, Gordon. (1987). *Autocracy.* Dordrecht: Kluwer Academic Publishers.

Wicksell, Johann and Knut, Gustav. (1896). *Finanztheoretische Untersuchungen nebst Darstellung und Kritik des Steuerwesens Schwedens.* Jena: Gustav Fischer.

INDEX